Negotiating international rules around pandemics p. 792

A strategy for solid-state cooling pp. 800 & 841

How geography and superfounders shaped a population p. 849

Science

$15
26 MAY 2023
science.org
AAAS



THE (UNSUNG) TONGUE

How a versatile organ amplified animal diversity p. 786

VISIT…

LANZAROTE Caliente.com

EDITORIAL

Lauer opens up

In March, *Science* published a news story by Jeffrey Mervis that chronicled five cases of individuals, mostly Chinese or of Chinese descent, whose research careers were disrupted or ended by personnel actions taken by the US National Institutes of Health (NIH). As I wrote in an accompanying editorial, these moves have eroded trust in the NIH and chilled important collaborations with China. The failure of the NIH to adequately explain and document publicly the reasons behind these actions has allowed the scientific community to assume the worst.

After the editorial and news story came out, Michael Lauer, the administrator most closely associated with these actions, invited me to the NIH for an off-the-record session, showing me the documentation behind many cases of individuals who have been affected by these investigations. (I agreed not to disclose any identifying information for individuals.) The cases generally followed a similar pattern in which individual scientists agreed to employment in a foreign country—mostly by so-called "talent" programs in China—but then did not disclose that employment to their university employer in the United States. Many of the faculty members in question also had grants from foreign countries that often were redundant with grants they received in the United States but had not disclosed this in NIH grant applications.

> **"If I'm sitting from the outside and watching this, I would quite frankly be quite scared…"**
> **Michael Lauer,** NIH

In this issue of *Science*, we are running a letter from Lauer and his NIH colleague Patricia Valdez that tells more of their story, much more than what Lauer told Mervis for the original article. Lauer also sat down with me for an on-the-record interview in which he was more forthcoming than he had been in the past. Anyone interested in these issues should read the full interview. Lauer agreed that because the NIH has been vague about the issue in the eyes of the scientific community, it was understandable that these actions were viewed with suspicion. "If I'm sitting from the outside and watching this, I would quite frankly be quite scared," he said, "especially when it's occurring in the backdrop of lots of other things that are happening in our world."

Lauer also agreed that the NIH could have done a better job of letting the community know that these actions were coming. Most of the communication that occurred was between the NIH and university administrators or scientific societies, but not with individual scientists. "That kind of conversation did not happen as well as it could," he said. "We didn't have the kind of conversations we should have had with on-the-ground scientists." As a result, individuals may not have realized that signing up for the talent programs was a breach of university and NIH policy and needed to be disclosed.

Lauer says he would have initiated a plan in which scientists suspected of violations involving international collaborations could have more easily come forward for clarification before an investigation ensued. In the small number of cases in which this has happened, much less disruptive resolutions have been achieved. In retrospect, Lauer acknowledged that such a system should have been set up years ago whereby violators were given the opportunity to come forward so that NIH could work with them and their universities to settle the matters administratively and to reach a common understanding about what the framework is for international collaboration. He said the fact that the Department of Justice was also carrying out parallel criminal actions tied his hands. That doesn't quite ring true because most of the cases did not involve criminal action. It's hard not to speculate that political factors were also at play.

The situation illustrates how the goals of scientists and politicians can result in conflict. Lauer's program has had bipartisan support in Congress, and many Americans outside the scientific community might see Lauer as doing important work to protect American interests. But the scientific community correctly sees the value of collaboration and support for scientists of all nationalities. The secrecy from the NIH up to this point has exacerbated this disconnect. In talking more openly and publishing their letter, the NIH is beginning to recognize the problem. Let's hope that continues.

–H. Holden Thorp



H. Holden Thorp
Editor-in-Chief, *Science* journals.
hthorp@aaas.org; @hholdenthorp

PHOTO: CAMERON DAVIDSON

10.1126/science.adi8239

NEWS

"If it's safe enough to drink, they should use it as drinking water."

South Korean opposition leader Lee Jae-myung, in the *South China Morning Post*, about Japan's plan to release treated radioactive water from the ruined Fukushima power plant into the ocean. Many countries neighboring Japan oppose the plan.

IN BRIEF

Edited by **Jeffrey Brainard**



A vaccine aims to protect infants from respiratory syncytial virus, a leading cause of infant hospitalization.

PUBLIC HEALTH

FDA advisers back maternal RSV shot

A panel advising the U.S. Food and Drug Administration (FDA) last week recommended that it approve a vaccine given to pregnant people to protect infants from respiratory syncytial virus (RSV), which can cause severe lung infections. The vote was unanimous based on the efficacy of the vaccine, called RSVpreF and branded Abrysvo. Ten members of the panel also endorsed the safety of the vaccine, which is designed to cause mothers to produce protective antibodies that their babies acquire during pregnancy. But four panel members weren't persuaded. A large, phase 3 trial by Pfizer, maker of the shot, found an elevated rate of premature births—5.7% in the vaccinated group versus 4.7% in the placebo group—but the difference did not reach statistical significance and neonatal deaths did not increase. Lower respiratory tract infections from RSV kill an estimated 46,000 babies younger than 7 months every year, hundreds of them in the United States, where RSV is the leading cause of infant hospitalization. The Pfizer vaccine was 69.4% efficacious in protecting this age group from severe disease. FDA is expected to rule in August whether to license the vaccine.

ARPA-H takes on osteoarthritis

FUNDING | The U.S. Advanced Research Projects Agency for Health (ARPA-H), a new federal funder charged with taking bold, innovative approaches to health research, last week announced its first program targeting a specific disease will focus on osteoarthritis. Thirty-two million people in the United States suffer from the degenerative condition, in which cartilage in the joints breaks down, causing pain and impairing mobility. Patients are typically treated with physical therapy and anti-inflammatory drugs and, when necessary, are offered metal joint replacements. ARPA-H's new program, Novel Innovations for Tissue Regeneration in Osteoarthritis, seeks to use a patient's own cells to regenerate lost bone and cartilage. ARPA-H was launched last year, modeled on the applications-focused, outside-the-box science sponsored by the Defense Advanced Research Projects Agency.

Cancer trials move goal posts

CLINICAL RESEARCH | Large clinical trials testing cancer therapies frequently alter their primary endpoint—the key health outcomes being measured—midstream, a study reports. A research team at the University of Texas MD Anderson Cancer Center and other institutions looked for design changes while the studies were underway by examining publicly available data from ClinicalTrials.gov, where trial sponsors post details about them, as well as in available protocol documents and publications reporting the studies. Of 755 phase 3 clinical trials, 145, or 19%, had such endpoint changes, including swapping the primary outcome measured for secondary ones, the researchers found; of those, 70% did not disclose the shifting endpoints in manuscripts, the team reported on 17 May in *JAMA Network Open*. The practice has raised concerns that researchers rework the endpoints to cast the trial results in a more positive light.

Mpox vaccine shows protection

INFECTIOUS DISEASES | A year after many countries started to immunize those at highest risk of mpox during a global outbreak, a study has shown the shots are effective



NEUROSCIENCE

Brain-spine link helps paralyzed man walk, navigate obstacles

A 40-year-old man whose lower body is partially paralyzed has been able to walk and navigate obstacles thanks to a digital bridge between his brain and spinal cord, researchers report this week in *Nature*. The international team had previously fitted Gert-Jan Oskam of the Netherlands with a stimulator that delivered electric pulses to his spinal cord, enabling him to walk over flat ground using crutches. But his movement was robotic, and he had difficulty navigating obstacles. In a first in a human, the team implanted electrodes above his motor cortex and connected them wirelessly via a headset to the stimulator. This allowed Oskam to walk more naturally and with more control. The team says it is working to test the technology in more patients and to make it less bulky.

A new brain-spine interface has allowed Gert-Jan Oskam, shown here working with a scientist, to enter and exit a car and stand at a bar while having a beer.

PHOTO: JIMMY RAVIER

against the monkeypox virus. The vaccine, called Jynneos and manufactured by Bavarian Nordic, was originally developed as a smallpox vaccine and licensed for mpox based largely on animal data. Now, researchers have used U.S. electronic health records to compare 2193 patients diagnosed with mpox with 8319 matched controls who were considered at high risk because they were living with HIV or taking pre-exposure prophylaxis to prevent HIV infection. (Mpox has primarily spread among men who have sex with men and their sexual networks.) Those in the control group were much more likely to have received the vaccine, the researchers report in *The New England Journal of Medicine*. They estimate it was 66% effective for those who received a full course of two doses and 35.8% for those who received only a single dose.

Proteins get $210 million gift

BIOMEDICINE | Immunologist Timothy Springer, a founder of vaccinemaker Moderna, this week announced he will give $210 million to a nonprofit research center he created to develop the use of proteins for medical research. The gift to the Institute for Protein Innovation in Boston is among the largest ever to a medical research center. It follows Springer's previous donations totaling $40 million to launch the institute in 2017 and expand it. The institute intends to provide scientists with synthetic antibodies and other protein tools to help illuminate fundamental biological processes and therapeutic leads. Springer, who still has a lab at Harvard Medical School, became a billionaire from his investments in research ventures including Moderna, which turned its expertise in messenger RNA into one of the most widely administered vaccines against SARS-CoV-2.

China vows ethics oversight

GENETICS | The Chinese Academy of Sciences (CAS) this month warned it will crack down on researchers over ethics violations, highlighting a case involving human embryonic development. A CAS ethics official told the academy's *China Science Daily* newspaper that investigators concluded researchers falsified an ethics review report for a study that produced cells resembling human embryonic stem cells in vitro and implanted chimeric embryos containing both human and mouse cells into female mice. CAS reduced the unidentified team leader's funding and suspended him from supervising postgraduates for a year, according to the news report. In an email to *Science*, Miguel Esteban, a stem cell biologist at CAS's Guangzhou Institutes of Biomedicine and Health, acknowledges he led the research in question, which he and colleagues published in *Nature* in March 2022. He denies falsifying documents and says the team followed international regulations and had ethical clearance for work on interspecies chimeras. In March, China's government announced revised rules for ethically problematic research involving human genetics, including requirements for ethics reviews. The mandate came 5 years after a Chinese scientist sparked worldwide outrage by announcing he had helped create genetically edited babies.

IN DEPTH



Conner Curran, 12, could climb stairs and walk better after receiving gene therapy for his Duchenne muscular dystrophy 5 years ago, but the effects are fading.

BIOMEDICINE

Gene therapy milestone looms, but field seeks better options

FDA is expected to endorse Duchenne muscular dystrophy treatment, as safety concerns linger over uses of viruses

By **Jocelyn Kaiser**

Five years ago, when Duchenne muscular dystrophy (DMD) began making it hard for him to walk, 7-year-old Conner Curran received a blood infusion of trillions of viruses carrying a muscle gene to replace his mutant one. Within 2 months the Connecticut boy went from crawling up stairs to "flying up," says his mother, Jessica Curran. The family and the researchers hoped he would never need another gene infusion. But the experimental treatment's effects are already fading.

Conner's story sums up the mixture of hope and disappointment among families and researchers as gene therapy for DMD, long seen as a potential cure for the debilitating and ultimately fatal disease, reaches a key milestone. This week, the U.S. Food and Drug Administration (FDA) was expected to approve a treatment similar to Conner's, developed by Sarepta Therapeutics.

But the therapy barely squeaked by an FDA advisory panel, with many members unconvinced that it works. Although "it's gratifying to see something approved after all these years, the current system is not perfect by any means," acknowledges muscular dystrophy researcher Jeffrey Chamberlain of the University of Washington, who helped lay the groundwork for the new gene therapies.

As a result, researchers are working on improvements, including strategies to give repeated doses to DMD patients like Conner and to edit the mutant disease gene with CRISPR instead of replacing it. They are also grappling with a major, unpredictable safety issue for some patients: toxicity from the high doses of the supposedly benign adeno-associated viruses (AAVs) used to deliver genetic treatments into muscle cells.

Last year, a young man with DMD died days after being given AAVs to ferry the DNA for CRISPR into his muscles and heart, and the virus has now been suspected or implicated in at least 11 previous gene therapy deaths, including a second DMD patient. "It's clear we've passed the maximally tolerated single dose" of AAV, says gene therapy researcher Barry Byrne of the University of Florida. Companies and labs are therefore trying to suppress the body's immune response to AAVs, which might decrease the risks of large doses or allow smaller, repeat doses. Immunosupression could also let previously treated patients "benefit from new, even better AAV therapies," says muscular dystrophy researcher Melissa Spencer of the University of California, Los Angeles (UCLA).

Because of a mutation in the gene for dystrophin, DMD patients lack functioning copies of the huge protein that serves as a shock absorber inside muscle fiber cells. Without it, muscle cells become damaged and gradually die. Patients usually end up using a wheelchair by age 12 and succumb to heart or respiratory problems by age 30. (Most are boys; the dystrophin gene is on the X chromosome, so girls have two copies and rarely develop DMD.) Existing therapies are only modestly effective.

The DNA encoding dystrophin is too large to package into the AAVs widely used in gene therapy. But inspired by an older man who was missing nearly half of the protein yet had only mild muscular dystrophy, Chamberlain's lab 2 decades ago devised a gene for a miniature dystrophin. Puppies with a version of DMD grew up with near normal muscle function when given this gene. Sarepta and three other companies, Pfizer (which treated Conner Curran), Solid Biosciences, and Généthon, went on to test delivering a micro- or minidystrophin gene into the muscles of young boys. Patients typically receive a one-time infusion of 1×10^{14} AAVs per kilogram of body weight, among the largest doses for any gene therapy.

At a 12 May FDA advisory meeting to consider Sarepta's request for approval of its gene therapy under an "accelerated" pathway, agency staff and outside advisers were skeptical largely because the company didn't have clear data confirming the intervention worked better than a placebo or that levels of microdystrophin in the

boy's muscles correlated with better muscle function. In the end, however, the advisory committee voted eight to six in favor of the therapy, which under an accelerated approval could be sold until a larger trial already underway is completed. If that study fails to definitively show benefits, the treatment could be pulled from the market. An FDA decision was expected by 29 May.

Even if the treatment proves to work, researchers now expect its effects to wear off. The AAV does not integrate the replacement gene into a cell's genome, instead delivering it into the nucleus as a loop of DNA. As damaged muscle fiber cells are being repaired by dividing muscle stem cells, some will not inherit the loop, and the benefits will fade as the modified cells become outnumbered.

Jessica Curran says this appears to be happening to Conner: Last fall, the sixth grader began using an electric scooter to conserve his strength in long school hallways. UCLA neurologist Perry Shieh, who is seeing declines in DMD patients he's treated in the Sarepta, Pfizer, and Solid trials, says, "The parents are asking: 'What is next for my child?'"

> **"This approval is going to help the field move forward."**
> **Melissa Spencer**, University of California, Los Angeles

For now, the answer is nothing because the boys make antibodies to AAV that would block any effort to retreat them the same way. But companies and academic researchers are trying to remove these antibodies with blood-filtering machines and drugs. They are also testing using other drugs to suppress immune cells that recognize AAVs or that make antibodies targeting the viruses.

"There are many overlapping, complementary strategies" that would allow AAV redosing, Byrne says. His group will soon launch a small study to see whether AAV antibodies can be lowered enough in Conner Curran and other DMD gene therapy patients to potentially allow them to be retreated. Although these immunosuppressive approaches carry their own risks, "compared to the consequences of disease … we feel it's justified," Byrne says.

Preventing a patient's immune system from making antibodies to an initial dose of AAV gene therapy might also make it possible to exchange a single large dose for smaller, repeat doses. An analysis of the October 2022 death of Terry Horgan, a 27-year-old DMD patient who received a custom-made CRISPR treatment designed to switch on a gene, underscored the need to reduce risks. Last week, a team of researchers, funded by the nonprofit his family had established to treat DMD and other diseases, posted a preprint on medRxiv absolving the gene editor and instead suggesting that the AAV he was given was toxic to his lungs and atrophied heart.

There's hope that CRISPR-based DMD gene therapies can eventually sidestep the AAV issue. Some teams are trying to use the same lipid nanoparticles, or fat bubbles, employed in the messenger RNA COVID-19 vaccines as a delivery vehicle for RNA encoding the gene editor's molecular components. But until researchers figure out how to steer the fat bubbles to muscle cells, AAV remains the only proven option for targeting the tissue.

Even if delivered by AAVs, CRISPR could still prove a better solution than current approaches that introduce a new dystrophin gene. For example, the gene editor could be used in some patients to "repair" the existing dystrophin gene in muscle cells by snipping out a sequence that causes cells to misread it. The gene would then produce nearly full-length dystrophin driven by natural promoters so it's made "at the right time, in the right place," notes molecular biologist Eric Olson of the University of Texas Southwestern Medical Center. And if the CRISPR therapy edits muscle stem cells, the changed gene should persist and its effects could be long-lasting.

Olson's approach, which his lab has demonstrated in mice and dogs, is under development by Vertex Pharmaceuticals. It hopes to start a clinical trial later this year. "The work is progressing well," Olson says.

CRISPR has its own downsides, however. The DNA-snipping Cas9 protein, one of its two components, comes from bacteria and can trigger an immune reaction against edited cells. As a result, "Your treated cells will be eliminated," says gene therapy researcher Dongsheng Duan of the University of Missouri School of Medicine, whose lab showed this phenomenon 2 years ago in dogs that got CRISPR for their DMD. Efforts are now underway to design Cas9 to be less immunogenic or to be quickly eliminated from cells after it makes the needed DNA cut.

For now, Spencer welcomes the expected approval of Sarepta's gene therapy. "We have to start somewhere. This approval is going to help move the field forward," she says. Eventually, we'll have improved therapies."

Jessica Curran hopes those improvements will come soon, and make it possible for her son to get another gene boost. "We have to figure out this antibody issue," she says. "Because these kids are all going to need [gene therapy] again." ■

COVID-19

New antibodies that the coronavirus can't elude

Researchers aim to make monoclonal antibodies that mutations in SARS-CoV-2 won't thwart

By **Robert F. Service**

In 2020, as the COVID-19 pandemic raged and other effective drugs were elusive, monoclonal antibodies (mAbs) emerged as a lifesaving treatment. But now, 3 years later, all the approvals for COVID-19–fighting antibodies have been rescinded in the United States, as mutations of the SARS-CoV-2 virus have left the drugs—which target parts of the original virus—ineffective.

Researchers around the globe are now trying to revive antibody treatments by redesigning them to take aim at targets that are less prone to mutation. "There are new approaches that present a much more challenging task for the virus to evade," says Paul Bieniasz, a virologist at Rockefeller University. Just this week, for example, researchers in Canada reported that they've created antibodylike compounds able to grab dozens of sites on viral proteins at the same time, acting as a sort of molecular Velcro to restrain the virus even if some of the sites have mutated to elude the drug candidate. Other researchers have taken less radical approaches to producing mutation-resistant antibodies.

All, however, worry that the work may be slow to reach the clinic. With the pandemic emergency declared over in the U.S. and other countries, governments and industry may have less incentive to develop promising new COVID-19 treatments. "There is no business model for this anymore," says Michael Osterholm, a public health expert at the University of Minnesota.

The antibodies that initially saved lives all glommed on to the tip of spike, the protein SARS-CoV-2 uses to attach to angiotensin-converting enzyme 2 (ACE2), a receptor on the surface of human cells. For the first 2 years of the pandemic, spike changed mod-

boy's muscles correlated with better muscle function. In the end, however, the advisory committee voted eight to six in favor of the therapy, which under an accelerated approval could be sold until a larger trial already underway is completed. If that study fails to definitively show benefits, the treatment could be pulled from the market. An FDA decision was expected by 29 May.

Even if the treatment proves to work, researchers now expect its effects to wear off. The AAV does not integrate the replacement gene into a cell's genome, instead delivering it into the nucleus as a loop of DNA. As damaged muscle fiber cells are being repaired by dividing muscle stem cells, some will not inherit the loop, and the benefits will fade as the modified cells become outnumbered.

Jessica Curran says this appears to be happening to Conner: Last fall, the sixth grader began using an electric scooter to conserve his strength in long school hallways. UCLA neurologist Perry Shieh, who is seeing declines in DMD patients he's treated in the Sarepta, Pfizer, and Solid trials, says, "The parents are asking: 'What is next for my child?'"

For now, the answer is nothing because the boys make antibodies to AAV that would block any effort to retreat them the same way. But companies and academic researchers are trying to remove these antibodies with blood-filtering machines and drugs. They are also testing using other drugs to suppress immune cells that recognize AAVs or that make antibodies targeting the viruses.

"There are many overlapping, complementary strategies" that would allow AAV redosing, Byrne says. His group will soon launch a small study to see whether AAV antibodies can be lowered enough in Conner Curran and other DMD gene therapy patients to potentially allow them to be retreated. Although these immunosuppressive approaches carry their own risks, "compared to the consequences of disease … we feel it's justified," Byrne says.

Preventing a patient's immune system from making antibodies to an initial dose of AAV gene therapy might also make it possible to exchange a single large dose for smaller, repeat doses. An analysis of the October 2022 death of Terry Horgan, a 27-year-old DMD patient who received a custom-made CRISPR treatment designed to switch on a gene, underscored the need to reduce risks. Last week, a team of researchers, funded by the nonprofit his family had established to treat DMD and other diseases, posted a preprint on medRxiv absolving the gene editor and instead suggesting that the AAV he was given was toxic to his lungs and atrophied heart.

There's hope that CRISPR-based DMD gene therapies can eventually sidestep the AAV issue. Some teams are trying to use the same lipid nanoparticles, or fat bubbles, employed in the messenger RNA COVID-19 vaccines as a delivery vehicle for RNA encoding the gene editor's molecular components. But until researchers figure out how to steer the fat bubbles to muscle cells, AAV remains the only proven option for targeting the tissue.

Even if delivered by AAVs, CRISPR could still prove a better solution than current approaches that introduce a new dystrophin gene. For example, the gene editor could be used in some patients to "repair" the existing dystrophin gene in muscle cells by snipping out a sequence that causes cells to misread it. The gene would then produce nearly full-length dystrophin driven by natural promoters so it's made "at the right time, in the right place," notes molecular biologist Eric Olson of the University of Texas Southwestern Medical Center. And if the CRISPR therapy edits muscle stem cells, the changed gene should persist and its effects could be long-lasting.

> **"This approval is going to help the field move forward."**
> **Melissa Spencer**, University of California, Los Angeles

Olson's approach, which his lab has demonstrated in mice and dogs, is under development by Vertex Pharmaceuticals. It hopes to start a clinical trial later this year. "The work is progressing well," Olson says.

CRISPR has its own downsides, however. The DNA-snipping Cas9 protein, one of its two components, comes from bacteria and can trigger an immune reaction against edited cells. As a result, "Your treated cells will be eliminated," says gene therapy researcher Dongsheng Duan of the University of Missouri School of Medicine, whose lab showed this phenomenon 2 years ago in dogs that got CRISPR for their DMD. Efforts are now underway to design Cas9 to be less immunogenic or to be quickly eliminated from cells after it makes the needed DNA cut.

For now, Spencer welcomes the expected approval of Sarepta's gene therapy. "We have to start somewhere. This approval is going to help move the field forward," she says. Eventually, we'll have improved therapies."

Jessica Curran hopes those improvements will come soon, and make it possible for her son to get another gene boost. "We have to figure out this antibody issue," she says. "Because these kids are all going to need [gene therapy] again." ■

COVID-19

New antibodies that the coronavirus can't elude

Researchers aim to make monoclonal antibodies that mutations in SARS-CoV-2 won't thwart

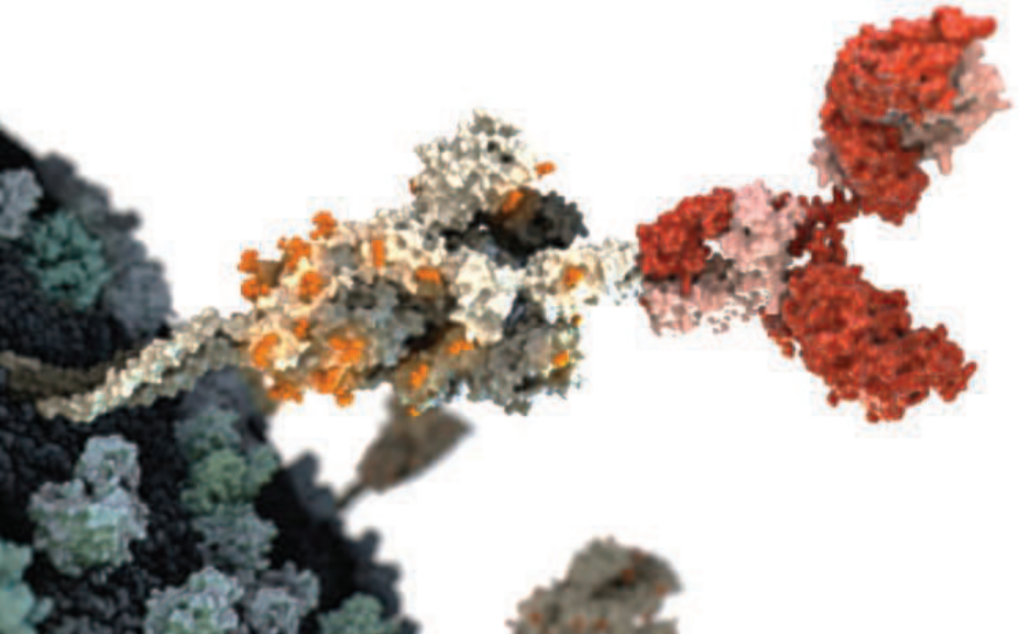
By **Robert F. Service**

In 2020, as the COVID-19 pandemic raged and other effective drugs were elusive, monoclonal antibodies (mAbs) emerged as a lifesaving treatment. But now, 3 years later, all the approvals for COVID-19–fighting antibodies have been rescinded in the United States, as mutations of the SARS-CoV-2 virus have left the drugs—which target parts of the original virus—ineffective.

Researchers around the globe are now trying to revive antibody treatments by redesigning them to take aim at targets that are less prone to mutation. "There are new approaches that present a much more challenging task for the virus to evade," says Paul Bieniasz, a virologist at Rockefeller University. Just this week, for example, researchers in Canada reported that they've created antibodylike compounds able to grab dozens of sites on viral proteins at the same time, acting as a sort of molecular Velcro to restrain the virus even if some of the sites have mutated to elude the drug candidate. Other researchers have taken less radical approaches to producing mutation-resistant antibodies.

All, however, worry that the work may be slow to reach the clinic. With the pandemic emergency declared over in the U.S. and other countries, governments and industry may have less incentive to develop promising new COVID-19 treatments. "There is no business model for this anymore," says Michael Osterholm, a public health expert at the University of Minnesota.

The antibodies that initially saved lives all glommed on to the tip of spike, the protein SARS-CoV-2 uses to attach to angiotensin-converting enzyme 2 (ACE2), a receptor on the surface of human cells. For the first 2 years of the pandemic, spike changed mod-



An antibody (right) bound to the surface spike protein of SARS-CoV-2 can block infection.

estly enough for the mAbs to continue to work. But as the virus encountered more people with antibodies from previous infections and vaccination, new variants emerged with extensive mutations in the ACE2-binding region, known as the receptor-binding domain (RBD). The variants dodged treatment and left pharma companies scrambling. "By the time you've isolated a good [mAb] the virus has moved on," says Laura Walker, who heads infectious disease biotherapeutics discovery and engineering for Moderna.

Now, researchers are seeking antibodies targeting segments of spike that the virus can't mutate without losing its ability to infect cells. "People are fishing for that hidden gem that targets something so conserved that the virus cannot mutate away from it," says Jean-Philippe Julien, an immunologist at the University of Toronto.

In March, for example, an international team led by researchers at the University of Italian Switzerland reported that it had isolated several human antibodies aimed at conserved targets on spike, unchanged across multiple viral variants. One binds to a site known as the fusion peptide, preventing the virus from merging with human cells. In cell-based assays, the antibody bound to four separate families of coronaviruses, including SARS-CoV-2. Another antibody, which targets a spike site known as the stem helix, blocked all SARS-CoV-2 variants from fusing with human cell membranes. The group, reporting in the 10 March issue of *Science Immunology*, also found that a "bispecific" antibody that binds to both the RBD and a separate region called subdomain 1 (SD1) that's involved in cell fusion protected mice against ancestral and Omicron SARS-CoV-2 variants.

Other groups are pursuing the same strategy. Researchers at the Fred Hutchinson Cancer Center (FHCC) reported in a March bioRxiv preprint that they, too, have isolated an SD1-targeting antibody that protects mice against all the recent variants of concern. And in January, a group led by antibody biologist Joshua Tan at the National Institute of Allergy and Infectious Diseases reported in *Cell Host & Microbe* that other fusion peptide and helix-binding antibodies can neutralize a broad array of SARS-CoV-2 variants in animals. "There are more and more antibodies that act more broadly because they are targeting spike in different areas," says Julie Overbaugh, a virologist at FHCC who led the SD1 work.

A separate approach takes aim at the human protein, ACE2, that SARS-CoV-2 and its relatives bind to on the cell surface. Last week, Bieniasz and his colleagues reported encouraging results in *Nature Microbiology*. They injected mice with copies of a soluble version of the human ACE2 receptor. Thirty-five days later, they screened the animals' blood serum for antibodies that targeted ACE2 and blocked SARS-CoV-2 from binding to it. They selected the most potent and injected it into mice that had been infected with a SARS-CoV-2 variant or a variety of other sarbecovirusus, the group of human and animal viruses that includes SARS-CoV-2. The antibody "was equally effective against all of them," Bieniasz says.

"This looks quite promising," Overbaugh says. But she and others are concerned that targeting human proteins could prompt side effects. They worry about interfering with ACE2's normal function, as it helps regulate blood pressure among other duties. Recent reports have added to the concern by suggesting that people with Long Covid may be producing antibodies against their own proteins, including ACE2 (*Science*, 28 January 2022, p. 364). Bieniasz agrees that more animal and human trials of the strategy will be needed, but he notes that in the initial cell culture studies, his group's antibody does not seem to keep ACE2 from working properly.

A third strategy looks to modify the structure of antibodies themselves in hopes of making them more potent. Antibodies are usually Y-shaped, with two arms that can attach to two separate targets. Julien and his colleagues have designed a family of spherical "multibodies," each with 24 attachment sites. In its most recent study, published this week in *Science Translational Medicine*, the Toronto team designed two different multibodies, one in which all 24 binding sites targeted the same site on SARS-CoV-2's spike protein, the other that targeted three different sites. When they injected their multibodies into infected mice, they found that both designs neutralized the virus at doses well below those needed for conventional antibodies. The three-target multibody also neutralized all recent subvariants and a wide array of viruses more distantly related to SARS-CoV-2.

Walker, who was impressed by these results, cautions that because the multibody strategy is new, it faces a longer road to the clinic. Researchers must verify that the multibodies remain in circulation for days—if not months—after infusion, and developers must show that the drugs can be manufactured reliably and cheaply. "It's not enough to have [mAbs] that will work, but [it's also] whether they will be available," Osterholm says.

But the momentum needed to turn the new antibodies into approved drugs may be waning. In March, President Joe Biden's administration launched Project Next Gen to help commercialize vaccines, mAbs, and other therapeutics. But the $5 billion for the effort could soon evaporate, a likely victim of ongoing negotiations between the administration and Congress over the U.S. debt ceiling. With little government help, pharma companies may be loath to pour hundreds of millions of dollars into commercializing new treatments. "It's going to take long-term investment," Osterholm says. "That is something we are missing."

At the same time, Tan notes, "There are at least some companies that remain interested." At least one sizable market remains—people who are immunocompromised, some 3% of the U.S. population. Still, pushing for new treatments is "definitely more challenging than it was a year or two ago," he says. ■

INFECTIOUS DISEASE

U.S. opens new high-security livestock laboratory

Unease greets Kansas facility that will work with dangerous agricultural pathogens

By **Warren Cornwall**

Virologist Robert Cross's lab is equipped to handle some of the world's most dangerous viruses. At the Galveston National Laboratory he has worked with guinea pigs infected with Ebola virus and macaques carrying Lassa fever. What it can't accommodate are pigs, which are common carriers of the deadly Nipah virus. "We're not really geared to handle large animals," says Cross, who wears a pressurized biosafety suit for his studies. "You can't just pick them up when you're wearing these space suits."

That's one reason why Cross is welcoming this week's ceremonial opening of a massive new high-security laboratory in Kansas, the first in the United States designed with pens and equipment to move big animals infected with the most hazardous agents, including Nipah virus. Although active research won't begin at the National Bio and Agro-Defense Facility (NBAF) in Manhattan for several years, Cross predicts that "a high containment resource to deal with agriculturally important pathogens … is going to change the [research] landscape."

The laboratory, which will be operated by the U.S. Department of Agriculture, has taken nearly a decade longer to complete than planned and, at $1.25 billion, cost nearly three times as much as first predicted. It is also controversial. Although many researchers and powerful Kansas politicians have supported the project, some scientists, Manhattan residents, and farm groups have voiced concerns about handling pathogens dangerous to livestock and humans in the nation's agricultural heartland. If a highly contagious animal disease escaped the lab, "it would just shut down commerce," says Larry Kendig, a board member of the Kansas Cattlemen's Association.

NBAF is intended to be the new home for work done for more than 60 years at the Plum Island Animal Disease Center, a federal laboratory on a tiny island in Long Island Sound, like an Alcatraz for diseases. Two decades ago Plum Island's aging facilities and lack of the tightest biosecurity features prompted federal officials to start to plan an upgrade.

When New York state politicians opposed handling even more dangerous viruses on the island, the U.S. Department of Homeland Security (DHS), the lab's overseer, looked elsewhere. It chose Manhattan, a town of 55,000 residents that is also home to Kansas State University. The agency cited broad community acceptance and the fact that Kansas was an emerging research hub for animal disease.

The new facility will have more than 53,000 square meters of floor space. Officials say staff will follow strict safety protocols and use specialized equipment, such as plastic "moon suits" with their own air supply. The features help make NBAF a biosecurity level-4 (BSL-4) lab, the most secure ranking. Plum Island, by comparison, is BSL-3.


The National Bio and Agro-Defense Facility is now formally open, but active research won't begin for a few more years.

NBAF also has equipment and rooms designed to handle large animals such as cows and pigs, enabling scientists to study diseases that are off-limits at Plum Island. That includes Nipah, a virus carried by fruit bats that can sicken pigs and people, killing between 40% and 75% of human patients. The routes of infection are mysterious. "We don't really understand what happens out on the farm in Asia where pigs are serving as amplifiers and spreading [the virus] to humans," says Lisa Hensley, an epidemiologist and virologist who heads the NBAF unit that will conduct research on zoonotic and emerging diseases.

Being able to safely work with pigs could help reveal how Nipah virus spreads and help researchers develop countermeasures. "I can see a Nipah vaccine … being a high interest item in coming years," says Cross, who is consulting with NBAF as it readies to open.

But some Kansans worry a pathogen could escape. "I don't understand why you'd put a facility like that in the middle of a highly productive cattle and crops area," says Kendig, whose ranch sits roughly 200 kilometers west of Manhattan.

The decision to put NBAF "in the middle of livestock country and tornado alley was real hubris," says Laura Kahn, a biodefense expert who worked at Princeton University's Program on Science and Global Security for much of the past 2 decades.

She points to a 2010 DHS evaluation that found the Kansas lab would have a 70% chance of triggering an outbreak of highly contagious foot and mouth disease over 50 years. A later review by a U.S. National Academies of Sciences, Engineering, and Medicine panel concluded that study was incomplete. But it also blasted a second DHS study that pegged the risk at only 0.11%, saying it relied on "questionable and inappropriate assumptions" that likely understated the risk.

Yet former senior biosecurity official Gerald Epstein warns there's also a danger to not having such a laboratory in an era of emerging infectious diseases. "I'm confident we know how to build this facility safely," says Epstein, who was deputy assistant secretary of homeland security for chemical, biological, radiological, and nuclear policy under former Presidents Barack Obama and Donald Trump. (From 2009 to 2012, Epstein also directed a science and security program at AAAS, the publisher of *Science*.)

Hensley ticks off the safety measures that will be in place, but also points to a more personal gauge of her confidence: She moved to Manhattan with her teenage son to help run the lab. "I literally live several miles from the lab," she says. ■

PHOTO: U.S. DEPARTMENT OF AGRICULTURE



Herring boat owners are challenging a rule requiring them to pay for onboard monitors.

POLICY

Legal challenge could weaken science's role in U.S. regulation

Supreme Court takes up industry case against fisheries rule

By **Jeffrey Mervis**

Fishing for Atlantic herring may seem worlds away from restrictions on power plant emissions or responses to the COVID-19 pandemic. But a case before the U.S. Supreme Court could affect all those activities and more by altering how federal agencies apply scientific expertise in carrying out their regulatory duties.

On 1 May the high court agreed to hear a case brought by four family-owned herring businesses. They argue that the National Marine Fisheries Service has no authority to require them to pay for onboard observers who would monitor their catch to help the agency protect declining fish stocks.

But legal scholars and environmentalists say much more than fishing is at stake in *Loper Bright Enterprises* v. *Raimondo*. When the high court rules, most likely in the spring of 2024, it may also take a step back from a long-standing legal doctrine, called the Chevron deference, used to determine whether government agencies have exceeded their authority. The doctrine has been applied for nearly 40 years to everything the government does, including regulations affecting businesses, schools, and the general population.

Abandoning or limiting Chevron could restrict an agency's ability to apply the best science in drawing up regulations, argues Julie McNamara of the Union of Concerned Scientists's climate and energy program. "We don't know how it will play out," says McNamara, whose organization is considering filing a brief urging the court to side with the government. But if Chevron is overturned, she says, "the chances of a successful court challenge to any government action designed to protect the public go way up."

Overturning Chevron is exactly what many conservative groups requested in briefs leading up to the high court's decision to take up the case. "Chevron deference is unconstitutional and ahistorical," wrote lawyers for the Cato Institute and the Liberty Justice Center. "It has wreaked havoc in the lower courts upon people and businesses."

The doctrine gets its name from a 1984 case, *Chevron U.S.A., Inc.* v. *Natural Resources Defense Council*, in which environmental groups sued then-President Ronald Reagan's administration over its interpretation of a provision of the 1977 Clean Air Act regulating emissions from power plants. The groups lost their challenge to what they saw as the administration's industry-friendly definition of a pollution source. But in unanimously upholding the administration's position, the justices also laid out a two-step process for deciding how federal courts should mediate such disputes in the future.

Courts have no cause to intervene if Congress has been clear about its intent, the high court explained. But if there are ambiguities or gaps in the law, they added, courts should defer to the agency's interpretation if it is reasonable and based on compelling evidence.

"The Supreme Court took the [Chevron] case in order to tell the lower courts to stop substituting their view of the right policy in places where the law seems ambiguous," says David Doniger, senior director for climate and energy policy at the National Resources Defense Council (NRDC), who argued the 1984 case. "Instead, it said in Chevron that you should leave that choice to the agency and then overrule them only if what they've done is totally wacko."

Sometimes Congress intentionally creates gray areas, Doniger explains, so that the law can be adapted to fit an ever-changing world. For example, the Clean Air Act does not list specific pollutants the federal government can regulate. Decades later, that lack of specificity allowed regulators to add carbon dioxide as part of the government's plan to fight climate change. In its Loper brief defending the use of the Chevron doctrine, the U.S. solicitor general argued it "respects the expertise agencies can bring to bear in administering complex statutory schemes."

Chevron has become "the most cited decision in all of American public law," says Columbia University law professor Thomas Merrill, who last year published a definitive history of its use by the courts. But the impact of applying the Chevron doctrine in any particular case is far from preordained.

"During the Reagan years, most of the conservatives and the business interests thought that the Chevron doctrine worked in their favor," Doniger says. "But during the Clinton and Obama years, agencies began to come up with more assertive interpretations of their authority in places where they found ambiguity. And the conservatives began to be concerned that Chevron, instead of being a restraint on regulation, was actually making these laws more powerful."

The new conservative majority on the high court has handed down a series of decisions carving out what are essentially exceptions to Chevron—situations in which the courts

need not defer to an agency's interpretation of a statute. The cases cover a broad range of government activities, including the scope of tax credits under the Affordable Care Act and requirements for COVID-19 vaccination or testing in the workplace.

The most prominent exception, known as the "major questions" doctrine, states that an agency can't carry out a major shift in government policy without explicit orders from Congress. "It says that, when the issues are sufficiently significant, politically controversial, and affect large segments of the public and substantial resources, the agency is not going to be given authority unless it can point to clear authorization in the statute," Merrill explains. Chief Justice John Roberts used the phrase in *West Virginia* v. *Environmental Protection Agency*, a landmark 2022 decision in which the court invalidated the agency's Clean Power Plan to reduce carbon dioxide emissions.

Despite those new carve-outs, court watchers say a ruling overturning Chevron would be a very big deal. "For 4 decades, Congress has written new statutes on the premise that this is how they will be interpreted, and lots of companies and organizations have ordered their affairs based on it," Doniger says.

Merrill believes the Supreme Court agreed to hear Loper because a rule that applies to herring boats is clearly not a major question. And he thinks the court will shy away from giving judges a free hand to block anything they see as government overreach.

"My guess is that the court will say that Chevron is still the framework for cases involving minor questions, rather than major questions," he says. "But it may also make a few revisions that move away from the maximalist interpretation of the doctrine, that whenever there's uncertainty in the law, the agency's view should prevail."

Even small changes in Chevron could have major consequences, Doniger says. Congress isn't capable of writing legislation that spells out exactly how every provision should be implemented, he argues. "The bandwidth isn't there, the expertise isn't there," he says. Even if it could, he adds, today's hyperpartisan political climate makes it difficult for Congress to pass any new major legislation.

NRDC hasn't decided whether to file a brief in support of the government. But Doniger offered this preview of what he might tell the high court. "Whatever decision that you announce to modify or replace Chevron, you've got to take into account the implications for having a government that can address modern problems," he says. "So, tread carefully." ■

NUCLEAR SECURITY

U.S. planning test reactor to run on weapons-grade uranium

Use of highly enriched fuel in civilian reactor would contravene decades-old nonproliferation policy

By **Adrian Cho**

The U.S. Department of Energy (DOE) is planning a small test reactor that would burn a large amount of weapons-grade uranium, according to the project's draft environmental assessment. The experiment, to be built in a cost-sharing arrangement, would provide data for a new type of power reactor being developed by TerraPower and Southern Company Services. But the use of highly enriched uranium, first reported by *Physics Today*, would contravene the U.S. policy of removing HEU from civilian reactors around the world to keep it from being made into bombs.

The decision is "discouraging," says Edwin Lyman, a physicist and director for nuclear safety at the Union of Concerned Scientists. "When the U.S. preaches the nonproliferation gospel, it should practice what it preaches." Alan Kuperman, a political scientist at the University of Texas at Austin, says, "There was not by any means adequate public disclosure by the department that they were planning to contradict 5 decades of U.S. nonproliferation policy."

Neither DOE nor Idaho National Laboratory (INL), where the test reactor will be built, would comment on the issue.

The Molten Chloride Reactor Experiment (MCRE) would differ dramatically from conventional power reactors. They consume uranium fuel enriched to roughly 4% uranium-235, the fissile isotope, and encased in metal rods. Some uranium atoms split or fission to release energy and neutrons, which then split other uranium atoms in a chain reaction. Pressurized water flows around the rods both to slow the neutrons so that they split atoms more effectively and to carry heat to steam generators that ultimately drive turbines to generate electricity.

The MCRE would instead be cooled by molten salt, into which the uranium would be dissolved. In theory, a molten salt reactor could burn used fuel from conventional reactors and generate less long-lived radioactive waste, Kuperman says. Because the salt would not slow the neutrons, the reactor would need fuel with higher enrichment, which would generate more neutrons.

TerraPower's commercial reactor would use fuel enriched to as much as 19% uranium-235, so-called high-assay, low-enriched fuel. But the MCRE will run on HEU enriched to greater than 90%—630 kilograms of it. That's hundreds of times more than some research reactors use and enough to make dozens of bombs, Kuperman estimates.

> **"This is going to be seen as hypocritical by many, many people."**
> **John Tierney**, Center for Arms Control and Non-Proliferation

The uranium is leftover from another research reactor that ran at INL from 1969 to 1990, he says.

Running on HEU should enable the MCRE to produce the data needed to design and license the molten-salt power reactor while remaining relatively small and inexpensive, Lyman says. DOE would cover $90 million of the MCRE's $113 million cost, and the reactor would start up in a few years. But its thrifty design would cost the United States credibility, says John Tierney, executive director of the Center for Arms Control and Non-Proliferation. "This is going to be seen as hypocritical by many, many people."

In the 1950s and '60s, the U.S. helped build research reactors around the world, providing HEU for many of them. In the 1970s, it changed course and led efforts to remove HEU from those reactors and repatriate it. Of the 171 research reactors that ran on HEU, 71 have switched to low-enriched fuel and 28 have shut down, according to the International Atomic Energy Agency—although five U.S. research reactors still use HEU.

The issue highlights a tension between DOE's Office of Nuclear Energy, which is eager to develop new reactors, and its National Nuclear Security Administration, which controls nuclear weapons and works for nonproliferation, Kuperman says. He and others have drafted a letter to DOE and President Joe Biden's administration to encourage them to reconsider the plan. "If they make the wrong decision, I think they're going to undermine much more of the nonproliferation regime than they realize." ■

need not defer to an agency's interpretation of a statute. The cases cover a broad range of government activities, including the scope of tax credits under the Affordable Care Act and requirements for COVID-19 vaccination or testing in the workplace.

The most prominent exception, known as the "major questions" doctrine, states that an agency can't carry out a major shift in government policy without explicit orders from Congress. "It says that, when the issues are sufficiently significant, politically controversial, and affect large segments of the public and substantial resources, the agency is not going to be given authority unless it can point to clear authorization in the statute," Merrill explains. Chief Justice John Roberts used the phrase in *West Virginia* v. *Environmental Protection Agency*, a landmark 2022 decision in which the court invalidated the agency's Clean Power Plan to reduce carbon dioxide emissions.

Despite those new carve-outs, court watchers say a ruling overturning Chevron would be a very big deal. "For 4 decades, Congress has written new statutes on the premise that this is how they will be interpreted, and lots of companies and organizations have ordered their affairs based on it," Doniger says.

Merrill believes the Supreme Court agreed to hear Loper because a rule that applies to herring boats is clearly not a major question. And he thinks the court will shy away from giving judges a free hand to block anything they see as government overreach.

"My guess is that the court will say that Chevron is still the framework for cases involving minor questions, rather than major questions," he says. "But it may also make a few revisions that move away from the maximalist interpretation of the doctrine, that whenever there's uncertainty in the law, the agency's view should prevail."

Even small changes in Chevron could have major consequences, Doniger says. Congress isn't capable of writing legislation that spells out exactly how every provision should be implemented, he argues. "The bandwidth isn't there, the expertise isn't there," he says. Even if it could, he adds, today's hyperpartisan political climate makes it difficult for Congress to pass any new major legislation.

NRDC hasn't decided whether to file a brief in support of the government. But Doniger offered this preview of what he might tell the high court. "Whatever decision that you announce to modify or replace Chevron, you've got to take into account the implications for having a government that can address modern problems," he says. "So, tread carefully." ■

NUCLEAR SECURITY

U.S. planning test reactor to run on weapons-grade uranium

Use of highly enriched fuel in civilian reactor would contravene decades-old nonproliferation policy

By **Adrian Cho**

The U.S. Department of Energy (DOE) is planning a small test reactor that would burn a large amount of weapons-grade uranium, according to the project's draft environmental assessment. The experiment, to be built in a cost-sharing arrangement, would provide data for a new type of power reactor being developed by TerraPower and Southern Company Services. But the use of highly enriched uranium, first reported by *Physics Today*, would contravene the U.S. policy of removing HEU from civilian reactors around the world to keep it from being made into bombs.

The decision is "discouraging," says Edwin Lyman, a physicist and director for nuclear safety at the Union of Concerned Scientists. "When the U.S. preaches the nonproliferation gospel, it should practice what it preaches." Alan Kuperman, a political scientist at the University of Texas at Austin, says, "There was not by any means adequate public disclosure by the department that they were planning to contradict 5 decades of U.S. nonproliferation policy."

> **"This is going to be seen as hypocritical by many, many people."**
> **John Tierney**, Center for Arms Control and Non-Proliferation

Neither DOE nor Idaho National Laboratory (INL), where the test reactor will be built, would comment on the issue.

The Molten Chloride Reactor Experiment (MCRE) would differ dramatically from conventional power reactors. They consume uranium fuel enriched to roughly 4% uranium-235, the fissile isotope, and encased in metal rods. Some uranium atoms split or fission to release energy and neutrons, which then split other uranium atoms in a chain reaction. Pressurized water flows around the rods both to slow the neutrons so that they split atoms more effectively and to carry heat to steam generators that ultimately drive turbines to generate electricity.

The MCRE would instead be cooled by molten salt, into which the uranium would be dissolved. In theory, a molten salt reactor could burn used fuel from conventional reactors and generate less long-lived radioactive waste, Kuperman says. Because the salt would not slow the neutrons, the reactor would need fuel with higher enrichment, which would generate more neutrons.

TerraPower's commercial reactor would use fuel enriched to as much as 19% uranium-235, so-called high-assay, low-enriched fuel. But the MCRE will run on HEU enriched to greater than 90%—630 kilograms of it. That's hundreds of times more than some research reactors use and enough to make dozens of bombs, Kuperman estimates. The uranium is leftover from another research reactor that ran at INL from 1969 to 1990, he says.

Running on HEU should enable the MCRE to produce the data needed to design and license the molten-salt power reactor while remaining relatively small and inexpensive, Lyman says. DOE would cover $90 million of the MCRE's $113 million cost, and the reactor would start up in a few years. But its thrifty design would cost the United States credibility, says John Tierney, executive director of the Center for Arms Control and Non-Proliferation. "This is going to be seen as hypocritical by many, many people."

In the 1950s and '60s, the U.S. helped build research reactors around the world, providing HEU for many of them. In the 1970s, it changed course and led efforts to remove HEU from those reactors and repatriate it. Of the 171 research reactors that ran on HEU, 71 have switched to low-enriched fuel and 28 have shut down, according to the International Atomic Energy Agency—although five U.S. research reactors still use HEU.

The issue highlights a tension between DOE's Office of Nuclear Energy, which is eager to develop new reactors, and its National Nuclear Security Administration, which controls nuclear weapons and works for nonproliferation, Kuperman says. He and others have drafted a letter to DOE and President Joe Biden's administration to encourage them to reconsider the plan. "If they make the wrong decision, I think they're going to undermine much more of the nonproliferation regime than they realize." ■

COVID-19

Decision nears for fall coronavirus vaccine

Consensus grows for abandoning the ancestral strain to improve immune responses

By **Jennifer Couzin-Frankel**

Earlier this year, U.S. regulators settled on a new strategy for COVID-19 vaccines. Like the annual flu shot, the vaccines will be updated each year based on the virus' evolution, then rolled out in the fall. Accordingly, on 15 June, advisers to the U.S. Food and Drug Administration will weigh which strain or strains of SARS-CoV-2 should make up the next iteration of vaccine, so that the agency can greenlight a version for companies to mass-produce.

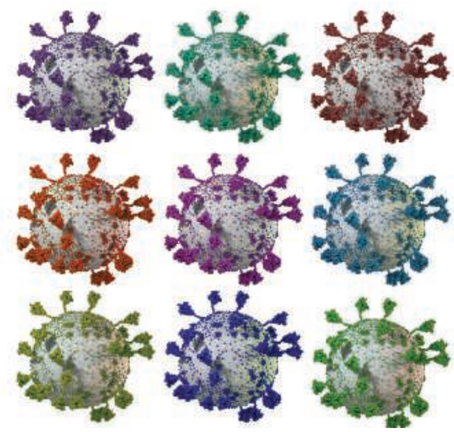
Regulators may well jettison the original SARS-CoV-2 strain that emerged in China and is long extinct—but which people are still being vaccinated against today. Many scientists favor eliminating it. The ancestral strain "should go out of the formulation," says William Messer, an infectious disease specialist and viral immunologist at Oregon Health & Science University. Last week, the World Health Organization (WHO) agreed. But other questions loom, including whether to bundle multiple virus strains into the vaccine or just one.

To date, COVID-19 vaccines have been modified only once, when a bivalent version based on both the original strain and the BA.5 Omicron variant was introduced in September 2022. Uptake was disappointing: Only 17% of people in the United States have rolled up their sleeves. (By comparison, about 50% get an annual flu shot.) Furthermore, many researchers say the bivalent vaccine packed less of a punch than it could have. The decision to preserve the ancestral strain sprang from worries that if an entirely new variant emerged, an Omicron-only vaccine might falter against it.

This hedging proved unwarranted: All major new variants have flowed from Omicron, which was first detected in South Africa in November 2021. And evidence increasingly shows that a vaccine split between a current strain and one that's extinct makes it harder for people to mount a strong immune response to the virus.

On 4 May, for example, David Ho, a virologist at Columbia University, and his colleagues posted a preprint study of 72 people, including some who had received four doses of the original vaccine and others who'd gotten three doses and a bivalent booster. Those who got the booster didn't produce antibodies that were notably better at neutralizing Omicron. The reason, Ho explains, is a phenomenon called immunological imprinting, in which repeatedly exposing the immune system to one strain—in this case, the ancestral one—skews the immune response in that direction. When the decision was made to keep the ancestral strain in COVID-19 vaccines, Ho says, imprinting "was probably not a dominant consideration, but it is now."

Florian Krammer, a virologist at the Icahn School of Medicine at Mount Sinai, agrees. He and his colleagues published a study this month in *The Lancet Microbe* in which they studied blood from 16 people 1 month before and about 2 weeks after they got a bivalent booster. After the booster, antibodies in the blood did a slightly better job of neutralizing the ancestral strain than BA.5. Krammer says his team also couldn't find "specific" antibodies solely reactive to BA.5, which could be especially protective if they're plentiful.



Regulators will decide soon which strains to include in a fall vaccine against the ever-changing coronavirus.

Last week, a WHO advisory group said in a statement that although current COVID-19 vaccines guard against severe disease, "protection against symptomatic disease is limited and less durable." In place of a bivalent shot, the group recommended a single-strain fall vaccine based on the XBB.1 lineage now dominating across continents, although it left the door open to other effective vaccine recipes.

Whether a single-strain XBB.1 vaccine is the best bet or whether multiple Omicron strains should be included is a point of debate. In the past few months, two closely related XBB substrains, XBB.1.5 and XBB.1.16, have crowded out other Omicron variants. "We're basically trying to guess what the next generation of variants will be, descending from which lineage," Ho says.

"From what we know now, matching the vaccine to whatever circulating variants you're trying to protect against probably does best," says Angela Branche, an infectious disease specialist at the University of Rochester. She co-leads a study called COVAIL that's examining immune responses spurred by different boosters. It has found that monovalent vaccines against Omicron perform somewhat better than those that include the ancestral strain.

An important question is whether vaccines better matched to current strains could reduce not just severe illness, but also transmission—something current vaccines appear to do poorly. An April study in *The New England Journal of Medicine* showed that after BA.5 faded and other Omicron strains surged, the bivalent vaccine's ability to prevent transmission peaked at about 30% 2 weeks after someone got the shot and fell to 0% at 16 weeks. "It's not an unreasonable supposition" that a closer match could perform a bit better, though the effect is still unlikely to persist, Messer says.

Some researchers also think the updated vaccines should not be limited to the messenger RNA formulations made by Pfizer and Moderna. Novavax makes a protein subunit vaccine, the technology used in hepatitis B and human papillomavirus vaccines. "It would be good to have protein vaccines for the fall," as those may give more durable protection, Branche says. But it's unclear whether the company would be able to mass-produce a new vaccine in time.

Robert Frenck, who directs the Vaccine Research Center at Cincinnati Children's Hospital and helped conduct trials of Pfizer's COVID-19 vaccine, points out that most vaccines for other infectious diseases "use one methodology," without causing concern. The strategy against COVID-19 need not be any different, he says.

Messer hopes regulators and companies will stay flexible as COVID-19 knowledge continues to grow—and urges targeting the new vaccines to people at highest risk. In the fall, "vaccine fatigue, COVID fatigue, is still going to persist," he says, and "triaging your efforts to get good vaccine uptake" will be vital. ■

FEATURES

TALES OF THE TONGUE

Since first evolving 350 million years ago, the tongue has taken myriad forms, unlocking new niches and boosting the diversity of life

By **Elizabeth Pennisi**



Mammalian tongues are very maneuverable thanks to an intricate network of muscle fibers.

Twice, quarterback Patrick Mahomes has led the Kansas City Chiefs to victory in the Super Bowl, the pinnacle of U.S. football. Although most fans have their eyes on the ball as Mahomes prepares to throw, his tongue does something just as interesting. Just as basketball star Michael Jordan did as he went up for a dunk, and dart players often do as they take aim for a bull's-eye, Mahomes prepares to pass by sticking out his tongue. That may be more than a silly quirk, some scientists say. Those tongue protrusions may improve the accuracy of his hand movements.

A small but growing group of researchers is fascinated by an organ we often take for granted. We rarely think about how agile our own tongue needs to be to form words or avoid being bitten while helping us taste and swallow food. But that's just the start of the tongue's versatility across the animal kingdom. Without tongues, few if any terrestrial vertebrates could exist. The first of their ancestors to slither out of the water some 400 million years ago found a buffet stocked with new types of foods, but it took a tongue to sample them. The range of foods available to these pioneers broadened as tongues diversified into new, specialized forms—and ultimately took on functions beyond eating.

"The incredible variation in vertebrate tongue form is replete with astonishing examples of almost unbelievable adaptation," says Kurt Schwenk, an evolutionary biologist at the University of Connecticut. Salamanders whipping out sticky tongues longer than their bodies to snag insects; snakes "smelling" their environment with their forked tongue tips; hummingbirds slurping nectar from deep inside flowers; bats clicking their tongues to echolocate—all show how tongues have enabled vertebrates to exploit every terrestrial nook and cranny. In humans, still more functions crowded aboard the tongue. "I am amazed by everything we do with our tongue: eat, talk, kiss. It's a central part of what it is to be a human," says Jessica Mark Welch, a microbial ecologist at the Forsyth Institute.

Managing these functions spurred the expansion of brain capacity, paving the way not just for throwing touchdown passes, but perhaps also for thinking on our feet. "The idea is that if you can reach with your tongue, you can reach with your hands, and you can reach with your thoughts," says Ian Whishaw, a neuroscientist at the University of Lethbridge. "Intuitively, perhaps we know this," he adds, when we use phrases like "tip of the tongue," "slip of the tongue," and "biting my tongue."

Yet how tongues came about "is one of the biggest mysteries in our evolutionary history," says Sam Van Wassenbergh, a functional morphologist at the University of Antwerp. Like other soft tissues, tongues are rarely preserved in fossils. Hidden inside the mouth, they defy easy observation. In the past decade, however, new technologies have begun to reveal tongues in action in different groups of animals. That work is beginning to yield new insights about the tongue's evolutionary trajectories, and how its specializations fueled further diversification. Kory Evans, an evolutionary biologist at Rice University, says the more biologists learn, the more convinced they are that "tongues are really fantastic."

A TONGUE TURNS OUT to be a slippery thing to define. Although tonguelike structures exist in virtually all vertebrates, from lampreys to mammals, "There is no clear definition to what makes a 'true tongue,'" says Daniel Schwarz, an evolutionary biologist at the State Museum of Natural History



Stuttgart. We tend to think of tongues as soft, muscular, and flexible—like our own. The human tongue is a muscular hydrostat, which, like a water balloon, must maintain the same overall volume when its shape changes. So, when Mahomes sticks out his tongue, it gets thinner overall than when it's just bunched up in his mouth; the same is true for a giraffe's purple tongue when it stretches 46 centimeters to snag leaves from a spiny tree branch.

But murkier cases exist elsewhere in the animal kingdom. The palatal organ of fish such as minnows, carp, and catfish can also be a bundle of muscle, but biologists are split on whether it should be considered a tongue. "Instead of being at the bottom of the mouth, it's at the top," says Patricia Hernandez, a functional morphologist at George Washington University. And despite many ideas, no one really knows this organ's function, Hernandez adds.

That's because fish don't need tongues like ours to swallow their food. They can rely on suction. They open their jaws wide, expand their throats, and pump water through their gill slits to create currents that sweep in food.

But, "The moment animals stick their head out of the water, suction becomes useless," says Schwenk, who has devoted his career to the study of animal tongues. Once those creatures made landfall, "they needed something to take the place of water" to draw prey into their gullet—and air is not dense enough. For millions of years, early landlubbers likely wriggled back to the ocean to swallow prey snagged on land. A few may have held their heads up high and let gravity do the work, like many birds today.

But the makings of a new way of feeding were already present in fish anatomy: a series of curved bones called branchial arches and the supporting muscles. In fish the branchial arches form the jaws, the hyoid bone that supports the back of the jaw, and the skeleton that forms the throat and gill slits. When fish feed, muscles supporting these structures generate suction by depressing and retracting the hyoid and expanding the gill slits to draw water in. To tongue specialists those motions look familiar. "The hyoid's movement to generate suction is very similar to movement of the tongue back and forth to manipulate prey," Schwenk explains.

Schwenk and Van Wassenbergh think that in early land vertebrates the branchial arches and related muscles began to change to form a "prototongue," perhaps a muscular pad attached to the hyoid that flapped when the hyoid moved. Over time, the pad became longer and more controllable, and more adept at grabbing and maneuvering prey (see graphic, p. 789).

Based on experiments with newts, Schwarz thinks a prototongue became functional even before the transition to land. Like other salamanders, newts are aquatic when young but mostly terrestrial as adults. Their metamorphosis, and the change in feeding strategies that accompanies it, might be akin to water-to-land changes that occurred hundreds of millions of years ago. And it holds a clue to how those changes might have unfolded.

Schwarz and his team found that before newts transform into full-fledged adults, they develop a tonguelike appendage that presses food against sharp, needlelike "teeth" on the roof of their mouth. The finding, which he and his colleagues reported in 2020, suggests a tonguelike structure may have helped early tetrapods feed, even before they climbed onto solid ground.

THE DEMANDS OF FEEDING may have prompted the emergence of the tongue, but natural selection then tailored and honed it for myriad other purposes, sometimes creating "ridiculously crazy specialized systems," Schwenk says. For example, web-toed salamanders (*Hydromantes*) whip out a sticky tongue to nab insects or other small arthropods, shooting their entire throat skeleton out through their mouth. This feeding mode involved retooling throat muscles, with one set storing elastic energy that could be instantaneously released to shoot out the tongue, and another set reeling the tongue back in.

Other salamanders, at least 7600 frogs and toads, as well as chameleons and other lizards have independently evolved other extreme forms of this quick-fire "ballistic" feeding. Chameleons, for example, launch their tongues at almost 5 meters per second, catching crickets in less than 1/10th of a second.

Ballistic feeding required adaptations in tongue surfaces and in the spit coating them. Copious gooey saliva exuded from barely visible protrusions called papillae can help make some frogs' tongues so sticky they can snare prey 50% heavier than themselves. The saliva coats the papillae, which can act like tiny sticky fingers to help grip prey, David Hu, a biomechanics researcher at the Georgia Institute of Technology, and his colleagues reported in 2017.

Horned lizards (*Phrynosoma*) use saliva-coated tongues not just to grab prey, but also to protect themselves from it. The ants they eat are powerful biters and particularly venomous, but the lizards swallow them alive. In 2008 Schwenk and Wade Sherbrooke, former director of the Southwest Research Station of the American Museum of Natural History, discovered that thick strings of mucus secreted by tongue and throat papillae incapacitate the noxious prey. More recently, Schwenk found that in horned lizards, the muscles that usually make up the sides of the tongue are only attached at the back. Evolution has reconfigured the muscles' free parts into ridges along the tongue's sides, possibly to create a mucous pocket for binding the ants before swallowing.

Whereas many frog and lizard tongues became fine-tuned for catching prey and getting it down the hatch, snake tongues instead evolved to provide an exquisite sense of smell, an adaptation that enables snakes to detect and sneak up on distant or hidden prey. Differences in the concentrations of an odorant sensed by each tine of a snake's forked tongue help the snake home in on quarry it can't see. As stereotyped as the tongue's flicking seems to be, it's actually quite malleable. Snakes that track prey both in water and in air, such as the northern water snake (*Nerodia sipedon*), modify their tongue's movements depending on whether their head is underwater, at the surface, or in the air, Schwenk and his former graduate student William Ryerson reported last year in *Integrative and Comparative Biology*. They seem to adjust the flicking pattern to optimize the collection of odor molecules in different conditions.

After studying the morphology, physiology, and tongue movements of dozens of reptile species, Schwenk is awed by how much they reveal about an animal's lifestyle. "If you just show me the tongue, I can tell you a huge amount," he says.

TONGUE EVOLUTION helped reptiles and amphibians capture animal prey, but in birds, some of the most outlandish tongue adaptations reflect a taste for plants. Most avian tongues are a stiff sliver of keratin (think fingernails) or bone, with little muscle or other living tissue. They "are just a conveyor belt to move food from front to back," Schwenk says. But there are exceptions—most notably in hummingbirds and other birds that feed on nectar. "The tongue is probably the most vital component for nectar feeding in birds," says David Cuban, a graduate student at the University of Washington (UW) who works with behavioral ecophysicist Alejandro Rico-Guevara.

Nectar is packed with energy and easy to find. But each flower offers just a drop or so, often sequestered in a long, narrow blossom. Many nectar-eating hummingbirds, sunbirds, and other unrelated groups of birds cope with these constraints by being small—



Like some other reptiles and many amphibians, this panther chameleon (*Furcifer pardalis*) shoots out its tongue to catch prey.

usually less than 20 grams—and having long slender bills and highly specialized tongues.

Researchers used to assume these birds relied on capillary action—the tendency of a liquid to flow up a narrow tube—to take in nectar. And some of them do, including the pied honeyeater (*Certhionyx variegatus*), Rico-Guevara's student Amanda Hewes and her collaborators have found. In this species the tongue has a paintbrush-like tip for picking up nectar, which is then drawn inward along grooves that run the length of the tongue.

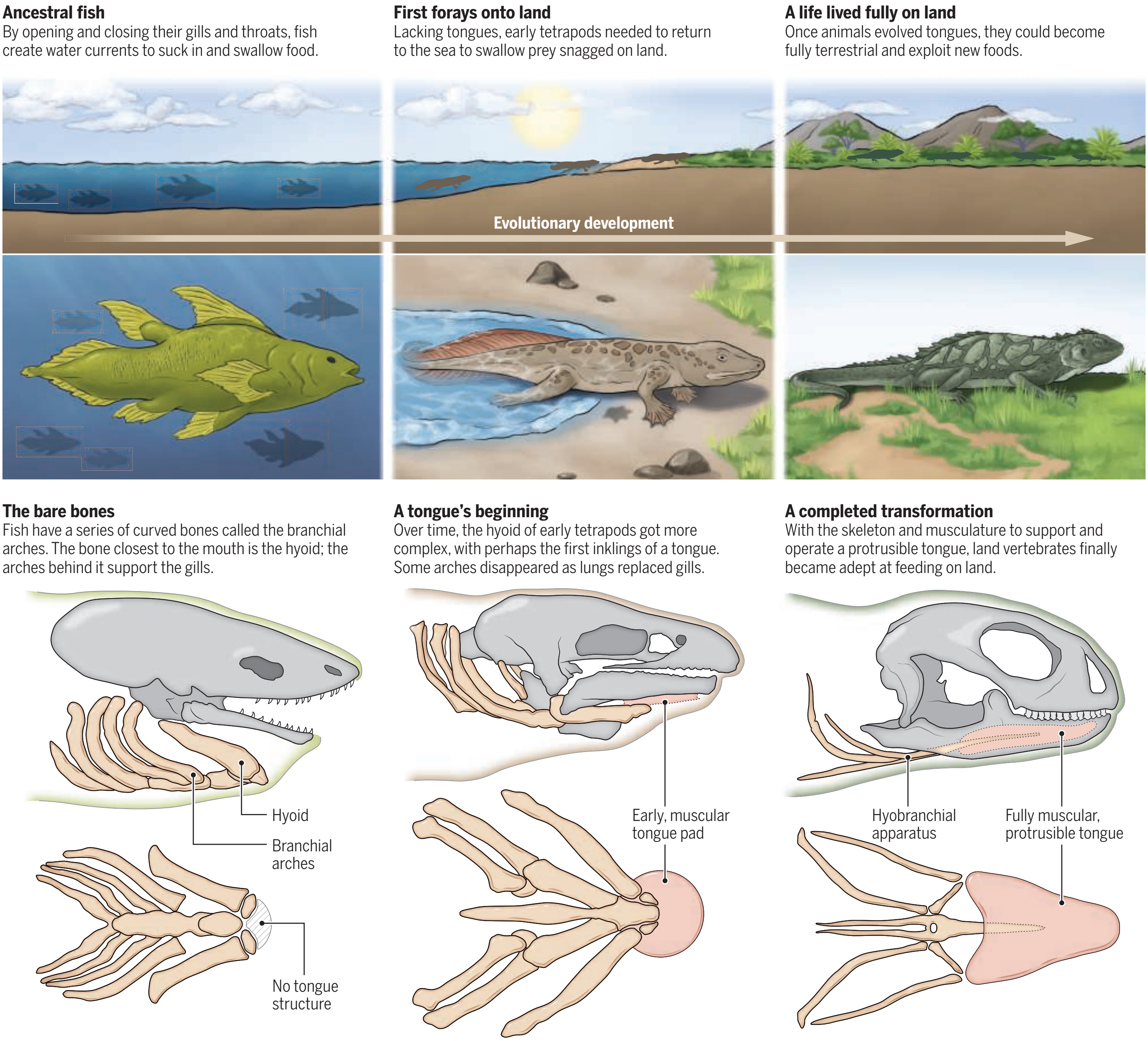
But for hummingbirds, which flick their tongues 15 times per second as they drain each flower and quickly move on, capillary action just isn't fast enough, Rico-Guevara says. His team captured high-speed videos as Anna's hummingbirds (*Calypte anna*), white-necked jacobins (*Florisuga mellivora*), sparkling violetears (*Colibri coruscans*), festive coquettes (*Lophornis chalybeus*), and other hummingbirds visited transparent artificial flowers loaded with artificial nectar. The movies revealed that the hummingbird tongue works like a tiny nectar pump.

Two grooves run from the tip about halfway back, lined with fringes that trap liquid. As the tip of the birds' flexible bill closes, it wrings nectar from fringes near the front of the tongue, pushing the liquid inward; then the bill opens at the base to help move nectar the rest of the way into the mouth, Rico-Guevara's team reported on 3 April in the *Journal of Experimental Biology*.

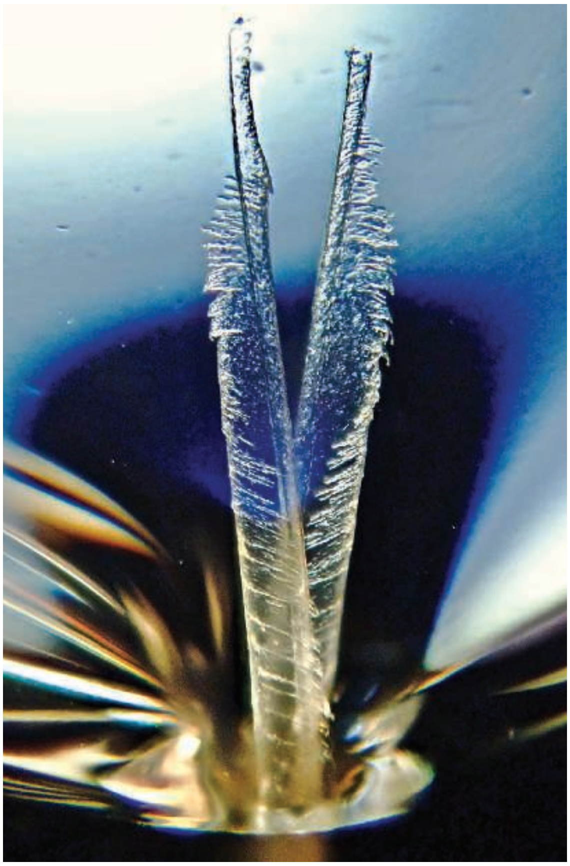
He and his collaborators have recently turned their attention to some of the oddest nectar-feeding birds: parrots. At

The dawn of the tongue

By making it possible to ingest food without suction, the evolution of the tongue some 350 million years ago was key to enabling vertebrates to move out of the sea and live on land. Skeletal structures originally used for opening gills had to evolve into the bones that could support a tongue and its movements.



GRAPHIC: A. FISHER/*SCIENCE*







Bird tongues have specialized in many ways to take advantage of different food sources. To sop up nectar, the tongue tip unfurls with fringes in Anna's hummingbird (left), and opens up with paintbrush-like bristles in lorikeets (top right). Green woodpeckers have barbs to harpoon insects (bottom right).

30 centimeters tall and 100 grams, the rainbow lorikeet towers over most nectarivorous birds and is utterly incapable of hovering in midair like a hummingbird. It has the typical short, stout, hooked parrot beak and a muscular tongue much like our own—all traits that make slurping nectar from long, thin blossoms impossible. But Rico-Guevara and Cuban have identified adaptations that enable these parrots to get the sweet stuff.

To start, the birds target flatter, more open blooms. And instead of hovering, they land on a nearby branch and contort their bodies around the flower. Then they open their beak and stick out their tongue, which undergoes an amazing transformation as it extends into a flower. The hard, scratchy tongue tip opens into a circular array of fine protrusions, Rico-Guevara recently discovered. "It looks like an anemone, almost," he says. These protrusions work like the bristles of a paintbrush to sop up nectar.

In one experiment, Rico-Guevara laced the test nectar solution with a barium compound, a diluted version of what doctors give patients to look for obstructions in the digestive tract, then took x-ray movies of lorikeet feeding. Once the tongue tip is saturated with a large drop of nectar, he found, the bird presses it against the top of the mouth, squeezing out the liquid. Then it closes its bill, nudging the nectar back toward the throat, and repeats the process until all the nectar goes down.

It's not the only way parrots consume nectar. Last year, Cuban filmed feeding in the more diminutive hanging parrots—so named because they sleep upside down. Instead of a bushy tongue tip like the lorikeet's, these parrots have a grooved tongue tip, and Cuban's videos reveal that they vibrate their tongues very quickly to pump tiny amounts of nectar back toward the esophagus and down the throat.

By describing in detail how these birds feed and calculating the energy they expend in the process, Cuban, Hewes, and Rico-Guevara hope to learn how their feeding strategies may have shaped their evolution—and that of the plants they feed on. Since evolving 22 million years ago, for example, hummingbirds have influenced how much nectar their partner plants produce and how deep their flowers are, and this in turn has influenced the length of the hummingbirds' beaks, their eagerness to monopolize flowers by chasing off competitors, and other traits. It's a coevolutionary dance of birds and flowers—mediated by their tongues.

IT'S IN MAMMALS, however, that the tongue displays its fullest versatility. The mammalian tongue has evolved into an intricate network of muscle fibers capable of moving in complex ways even without any bones, tendons, or joints. It contributes to suckling in most species, helps with thermoregulation in some (picture a panting dog), and takes on even more specialized tasks in a few, such as producing the sounds used for echolocation in bats and speech in humans. And it hosts the taste buds that help guide feeding in all these species. "The tongues of most mammals perform great feats," Hu says. "It's truly a multifunctional tool, and has only received less attention because it is less accessible than an animal's external appendages."

The tongue's most essential job in mammals is to position food to be chewed and swallowed. Depending on the species, that could mean shifting the food from one side to another with each bite or confining it to just one side, while the tongue itself stays safely away from chomping teeth. Then, with the addition of saliva it helps produce, the tongue shapes mashed food into a rounded "bolus" that can fit easily down the throat. Finally, it pushes that bolus back to be swallowed, making sure no food en-

ters the airways. In a sense, the tongue has become a "hand of the mouth," says J.D. Laurence-Chasen, a biologist at the National Renewable Energy Laboratory.

All this processing enables mammals to digest food more rapidly and efficiently, so they get more from their diet than most other animals. That bounty has fueled other evolutionary advances, such as high metabolic rate and activity, prolonged pregnancies, and large brains. Indeed, Callum Ross, a biomechanist and neurobiologist at the University of Chicago, counts the origin of mastication as one the three course-changing evolutionary transitions enabled by the tongue, along with the shift from water to land and the origin of human speech.

Until recently, researchers couldn't get a detailed view of how the tongue maneuvers food because lips, cheeks, and teeth got in the way. But lately Ross's group has been using a technique called x-ray reconstruction of moving morphology (XROMM) that involves recording the movements of surgically implanted beads with x-rays and turning the results into 3D animations.

In their experiments with opossums and monkeys, cameras simultaneously capture images from different angles as an animal eats or drinks, and the reconstructed animation allows the researchers to see how the tongue moves in relation to the jaws and teeth. "We are able to see features of movement that were utterly hidden," explains Elizabeth Brainerd, a functional morphologist at Brown University and an XROMM pioneer who has advised Ross on how to adapt this technology for his studies. By comparing tongue movements in different species, researchers hope to learn how tongue specializations may have contributed to the evolution of each animal's lifestyle and food preferences.

More recently, Laurence-Chasen and Ross worked with Chicago colleague Nicho Hatsopoulos and Fritzie Arce-McShane, now a neurobiologist at UW, to combine XROMM analysis with recordings of neural activity in monkeys. Such studies, they hope, will reveal how the brain coordinates the complex tongue movements involved in feeding, drinking, and perhaps even vocalizations. In one experiment, an array of electrodes monitored a penny-size region of cortex located behind the temple as monkeys munched on grapes. This region contains both sensory neurons that receive input from the tongue and mouth and motor neurons that send signals back to help control tongue movement. The team found that the firing pattern of the motor neurons accurately predicted the tongue's shape changes, they will report soon in *Nature Communications*.

The work upends the once-prevalent notion that chewing, like walking, is mainly under the control of the brainstem. The cortex is very much involved as well, ensuring that the tongue "is capable of complex, asymmetrical deformations" that adjust on the fly to gummy bears, steak, even milkshakes, Laurence-Chasen explains.

Whishaw wonders whether the human tongue's dexterity could have helped pave the way for our fine control of our hands and even our mind. His curiosity was piqued by an unexpected finding a few years ago. His team had taught mice to use their paws instead of their mouths to pick up fruit. They noticed that some animals stuck out their tongues as they reached with their paws, they reported in 2018.

In follow-up studies that have yet to be published, he, Duke University neurobiologist Xu An, and their colleagues have identified what they call the "oromanual" region of the cortex, a previously uncharted area that exerts control over both the hand and tongue. Whishaw thinks a similar brain region exists in humans and could help explain why so many people gesture as they talk, why children learning to write often twist their tongues as their fingers shape letters—a phenomenon noted by Charles Darwin—and even why Mahomes sticks his tongue out before a pass. He suspects many people move their tongue as they are about to use their hands—but because their mouth stays closed, no one is the wiser.

A common brain region for the hand and tongue makes evolutionary sense, Whishaw says. In early land animals, a dexterous tongue was essential for feeding; later, when some animals began grabbing food with their limbs, evolution might have coopted the same brain circuitry guiding the tongue to coordinate hand movements. He speculates that even more complex behaviors—such as thinking—could have arisen from the brainpower that initially evolved to coordinate the tongue. "I think it is the center of our being, as crazy as that might seem." ■

A home for microbes

The human tongue hosts a complex community of bacteria that can influence our health. "It's an unrecognized and really important part of the human microbiome," says Jessica Mark Welch, a microbial ecologist at the Forsyth Institute. Her team has developed a technique for labeling many of the more abundant bacteria while keeping the microbial community intact, allowing the researchers to map where each species resides on the tongue. Proportions of these microbes vary from person to person, Mark Welch says, but each may have a job. *Rothia mucilaginosa* (●), *Actinomyces* (●), *Neisseriaceae* (●), and *Veillonella* (●) convert nitrate to nitrite—something the human body cannot do—making nitrite available to help regulate blood pressure. Others may help prevent cavities or aid the immune system. "We don't know yet!" Mark Welch says. But seeing what's there is a first step toward finding out. —*E.P.*



PHOTO: STEVEN WILBERT AND GARY G. BORISY/FORSYTH INSTITUTE/CC BY NC ND

PERSPECTIVES

CELL BIOLOGY

Proteasome cap particle regulates synapses

Deubiquitylation by free 19S proteasome cap particle modulates synaptic transmission

By **Fulya Türker**[1] *and* **Seth S. Margolis**[1,2]

For 48 years, the 76–amino acid protein ubiquitin (*1*) has been studied and found to be vital in all aspects of mammalian biology, including the formation and function of the nervous system (*2*). Ubiquitin is conjugated to most cellular proteins in part to mediate adenosine triphosphate (ATP)–dependent protein degradation (*3*, *4*). This process entails the ubiquitylated proteins being targeted to the 19S regulatory cap on the 26S proteasome complex, where substrates are deubiquitylated, unfolded, and degraded through the 20S core particle (*5*). On page 811 of this issue, Sun *et al.* (*6*) report that there are soluble 19S particles free of the proteasome throughout rat cortical neurons. These autonomous 19S particles retain deubiquitylating activity and edit ubiquitylation of critical neuronal proteins, which modulates synaptic transmission.

Ubiquitylation of proteins is mediated by three enzymes (E1, E2, and E3) that target and place ubiquitin on the lysine (K) residues of a target protein. One of ~600 E3 enzymes places ubiquitin on a target protein in a variety of topologies, leading to distinct regulatory consequences. Some E3 ubiquitin ligases add a single ubiquitin to modulate protein-protein interactions, protein localization, and/or protein activity. By contrast, many E3 ubiquitin ligases add ubiquitin chains by linking multiple ubiquitin proteins on either the N-terminal methionine of ubiquitin or any one of seven lysine residues of the ubiquitin (*7*). These ubiquitin chains have a variety of regulatory outcomes, such as targeting for lysosomal or proteasomal degradation. The complex nature of this ubiquitin coding on proteins and the downstream signaling that ensues are areas of active research (*7*).

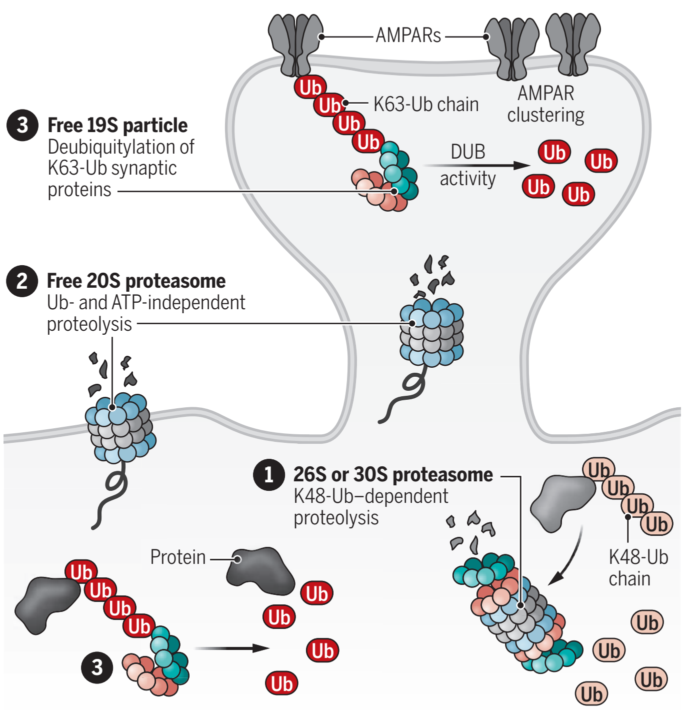
Once a polyubiquitylated protein has reached the 19S-capped 20S proteasome, the ubiquitin chains are edited by 19S regulatory particle–associated deubiquitylases (DUBs). 19S-dependent ubiquitin chain editing may help to secure a substrate's fate to be degraded or alter its fate by changing the pattern of ubiquitylation (*7*, *8*). Although a fraction of client proteins targeted to the 19S-capped proteasome are not degraded, substrates that are to be degraded become fully engaged with the 19S cap and are unfolded in an ATP-dependent manner concurrently with deubiquitylation. The unfolded protein is threaded through the narrow opening of the proteasome, and the catalytic residues of the 20S core particle degrade the substrate into peptides.

Ubiquitin-dependent degradation is crucial for the proper function of the nervous system. For example, inhibition of the proteasome by chemical antagonists (*9*) affects synaptic remodeling (*10*), physiology (*11*), and cell survival (*12*), and finely tuned proteasome function is required for higher brain functions, such as learning and memory (*2*, *13*), as shown in rodents. Additionally, human studies have shown a link between dysregulation of proteasome function and brain disease (*12*), including neurodegeneration. The 19S-capped 20S proteasome at neuronal synapses mediates local degradation. Moreover, a neuronal-specific 20S proteasome that is free of a 19S subunit functions at the neuronal plasma membrane to degrade intracellular proteins into extracellular signaling peptides (*14*). Sun *et al.* also observed free 20S proteasomes in the dendrites and synapses and free 19S particles near synapses in rat cortical neurons (see the figure).

Considering that the free 19S subcomplex cannot drive proteolytic function, Sun *et al.* investigated the possibility of it acting as a DUB. Upon examination of ubiquitin chains associated with the free 19S subunit, they showed that the free 19S particle is responsible for deubiquitylating K63-linked ubiquitin chains through the activity of the DUB ubiquitin C-terminal hydrolase isozyme L5 (UCHL5, also known as UCH37). Substrates of this free 19S particle DUB included many essential presynaptic and

Distinct proteasome subcomplexes in neurons

19S-capped proteasomes (26S and 30S) (**1**), free 20S proteasomes (**2**), and free 19S regulatory cap particles (**3**) are localized to dendrites and synapses in rodent cortical neurons. 26S and 30S proteasomes mediate adenosine triphosphate (ATP)– and Lys[48] (K48)–linked ubiquitin (Ub) chain–dependent protein degradation to modulate postsynaptic protein composition, leading to changes in synaptic remodeling. Free 20S proteasomes are shown mediating protein degradation of proteins in an ATP- and Ub-independent manner, contributing to intracellular protein turnover and production of extracellular signaling peptides. Free 19S regulatory subunits have deubiquitylase (DUB) activity toward synaptic proteins with K63-linked Ub chains, such as AMPA receptors (AMPARs), which alters neurotransmission.



[1]Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. [2]Solomon H. Snyder Department of Neuroscience, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. Email: smargol7@jhmi.edu

GRAPHIC: K. HOLOSKI/*SCIENCE*

postsynaptic proteins. Further research may reveal whether this DUB activity is relevant to pathway regulation in other cell types.

Sun *et al.* showed that inhibition of 19S DUBs in cultured rat cortical neurons with a selective inhibitor, b-AP15, led to an increase in miniature excitatory postsynaptic current (mEPSC) frequency—suggesting a negative regulation of synaptic vesicle release—and a decrease in mEPSC amplitude, an indicator of postsynaptic responsiveness. These results indicate that the free 19S subunit can alter synaptic transmission by regulating both pre- and postsynaptic compartments. Moreover, they identified the AMPA receptor (AMPAR), one of the main excitatory glutamate receptors on postsynaptic neuronal membranes, as a substrate for the 19S DUBs. Inhibition of DUBs led to a decrease in the total dendritic AMPAR pool, suggesting that the 19S DUBs stabilize and protect the receptors from degradation. Notably, Sun *et al.* also found that the free 19S subunit is a major regulator of AMPAR function by modulating its clustering at the plasma membrane. These findings of an independent function of 19S particles will spark further investigation of proteasome subcomplexes in the nervous system.

E3 ubiquitin ligases control substrate ubiquitylation upstream of 19S subunit regulation. Roughly 13% of known E3 ubiquitin ligases are mutated in neurological diseases (*15*). Furthermore, nearly a dozen E3 ubiquitin ligases have been found to be mutated in autism spectrum disorders. Interestingly, not all disease-associated E3 ubiquitin ligases mediate proteasome targeting. Moreover, many studies looking at E3 ubiquitin ligase deficits in mice observe little change in protein turnover. Taken together, these findings indicate the possibility that a ubiquitin code may be highly relevant to a protein's functionality, and deficits in the ubiquitylation or deubiquitylation pathways could be highly impactful to protein regulation and brain function, such as synaptic transmission, as demonstrated by Sun *et al.* ■

REFERENCES AND NOTES

1. G. Goldstein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 11 (1975).
2. F. Türker *et al.*, *Cell Chem. Biol.* **28**, 903 (2021).
3. J. D. Etlinger, A. L. Goldberg, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 54 (1977).
4. A. Hershko *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3107 (1979).
5. O. Coux *et al.*, *Annu. Rev. Biochem*. **65**, 801 (1996).
6. C. Sun *et al.*, *Science* **380**, eadf2018 (2023).
7. R. Yau, M. Rape, *Nat. Cell Biol*. **18**, 579 (2016).
8. D. Finley, *Annu. Rev. Biochem*. **78**, 477 (2009).
9. A. L. Goldberg, *J. Cell Biol.* **199**, 583 (2012).
10. M. D. Ehlers, *Nat. Neurosci*. **6**, 231 (2003).
11. B. Bingol, E. M. Schuman, *Nature* **441**, 1144 (2006).
12. M. Schmidt, D. Finley, *Biochim. Biophys. Acta Mol. Cell Res*. **1843**, 13 (2014).
13. R. Fonseca *et al.*, *Neuron* **52**, 239 (2006).
14. K. V. Ramachandran, S. S. Margolis, *Nat. Struct. Mol. Biol*. **24**, 419 (2017).
15. A. J. George *et al.*, *Front. Genet*. **9**, 29 (2018).

10.1126/science.adi1166

MEDICINE

Emerging frontiers in regenerative medicine

Bridging knowledge gaps could enable regenerative therapy

By **Kara L. McKinley**[1,2], **Michael T. Longaker**[3,4], **Shruti Naik**[5,6,7,8]

Nearly every human malady, be it injury, infection, chronic disease, or degenerative disease, damages tissues (*1*). Moreover, 45% of all deaths can be traced to inflammation- and fibrosis-related regenerative failures (*1*). Restoring health after damage requires the answer to a key question: How can human tissues be coaxed to regenerate? Identifying instructive cues that direct refractory tissues down a regenerative path remains a critical yet elusive goal. Nonetheless, approaches to target roadblocks that impede regeneration, including insufficient and/or functionally inadequate progenitor cells, fibrosis, and chronic inflammation, are continuing to progress from bench to bedside. Pivotal advances have been made to overcome these hurdles using cell therapy, in vivo reprogramming, synthetic biology, and antifibrotic and anti-inflammatory therapies, but many challenges remain and knowledge gaps must be addressed to make regeneration a mainstay of modern medicine (*1*).

The most conspicuous requirement for regenerative therapies is to replace the components of tissues that were lost or compromised by disease. Invigorating endogenous stem cells is an appealing strategy, but, to date, the greatest benefits have emerged from cell therapies. Adult stem cell–based regenerative therapies have shown clinical benefit to treat hematological malignancies, burn wounds, and ocular degeneration. These therapeutic modalities also lend themselves to gene editing to correct monogenic perturbations. For instance, a young boy suffering from junctional epidermolysis bullosa, a lethal skin disease caused by mutations in the laminin-*332* gene, was treated with an autologous (patient's own) skin transplant (*2*). A millimeter-sized sample of the boy's skin was collected and transduced with a retroviral vector expressing the wild-type laminin 332 cDNA; the tissue was then expanded ex vivo before being grafted to restore 80% of the body surface area.

Human pluripotent stem cell (hPSC)–based therapies have also shown promise and have entered clinical trials in the United States for type 1 diabetes (T1D; clinical trial NCT04786262), Parkinson's disease (PD; NCT02452723 and NCT03119636), and age-related macular degeneration (AMD; NCT04339764) (*3*). These three diseases are particularly amenable to stem cell–based therapies because they are associated with deficiency of a defined cell type: insulin producing pancreatic islets in T1D, midbrain dopaminergic (mDA) neurons in PD, and retinal pigment epithelial cells in AMD. Moreover, transplantation sites are surgically accessible. Lab-generated hPSCs are being assessed for safety and efficacy in reducing disease symptoms. A case report of a PD patient who received midbrain transplant of autologous hPSC-derived mDA neurons showed stable grafting without immune reaction and exhibited improved motor function 2 years after implantation (*4*).

Despite these early glimpses of success, cell therapies are hampered by many biological and technical hurdles. Autologous hPSC-based therapies derived from induced pluripotent stem cells (iPSCs) avoid immune aggravation, but this cost- and labor-intensive strategy requires safety testing for each use. A major concern with nonautologous cell therapy is immune rejection of the graft. Thus, "off-the-shelf" allogeneic therapies must be coupled to strategies that allow them to avoid rejection, such as systemic immunosuppression in T1D hPSC therapy (NCT04786262). An alternative strategy is to encapsulate the graft within a selectively permeable barrier to allow for diffusion of nutrients and the therapeutic molecules (e.g., insulin) while excluding immune components that would reject the engrafted cells. To avoid fibrotic foreign

[1]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [2]Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA. [3]Division of Plastic and Reconstructive Surgery, Department of Surgery, Standford University School of Medicine, Stanford, CA, USA. [4]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. [5]Department of Pathology, New York University Langone Health, New York, NY, USA. [6]Department of Medicine, New York University Langone Health, New York, NY, USA. [7]Ronald O. Perelman Department of Dermatology, New York University Langone Health, New York, NY, USA. [8]Perlmutter Cancer Center, New York University Langone Health, New York, NY, USA. Email: kara_mckinley@harvard.edu; longaker@stanford.edu; shruti.naik@nyulangone.org

postsynaptic proteins. Further research may reveal whether this DUB activity is relevant to pathway regulation in other cell types.

Sun *et al.* showed that inhibition of 19S DUBs in cultured rat cortical neurons with a selective inhibitor, b-AP15, led to an increase in miniature excitatory postsynaptic current (mEPSC) frequency—suggesting a negative regulation of synaptic vesicle release—and a decrease in mEPSC amplitude, an indicator of postsynaptic responsiveness. These results indicate that the free 19S subunit can alter synaptic transmission by regulating both pre- and postsynaptic compartments. Moreover, they identified the AMPA receptor (AMPAR), one of the main excitatory glutamate receptors on postsynaptic neuronal membranes, as a substrate for the 19S DUBs. Inhibition of DUBs led to a decrease in the total dendritic AMPAR pool, suggesting that the 19S DUBs stabilize and protect the receptors from degradation. Notably, Sun *et al.* also found that the free 19S subunit is a major regulator of AMPAR function by modulating its clustering at the plasma membrane. These findings of an independent function of 19S particles will spark further investigation of proteasome subcomplexes in the nervous system.

E3 ubiquitin ligases control substrate ubiquitylation upstream of 19S subunit regulation. Roughly 13% of known E3 ubiquitin ligases are mutated in neurological diseases (*15*). Furthermore, nearly a dozen E3 ubiquitin ligases have been found to be mutated in autism spectrum disorders. Interestingly, not all disease-associated E3 ubiquitin ligases mediate proteasome targeting. Moreover, many studies looking at E3 ubiquitin ligase deficits in mice observe little change in protein turnover. Taken together, these findings indicate the possibility that a ubiquitin code may be highly relevant to a protein's functionality, and deficits in the ubiquitylation or deubiquitylation pathways could be highly impactful to protein regulation and brain function, such as synaptic transmission, as demonstrated by Sun *et al.* ■

REFERENCES AND NOTES

1. G. Goldstein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 11 (1975).
2. F. Türker *et al.*, *Cell Chem. Biol.* **28**, 903 (2021).
3. J. D. Etlinger, A. L. Goldberg, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 54 (1977).
4. A. Hershko *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3107 (1979).
5. O. Coux *et al.*, *Annu. Rev. Biochem.* **65**, 801 (1996).
6. C. Sun *et al.*, *Science* **380**, eadf2018 (2023).
7. R. Yau, M. Rape, *Nat. Cell Biol.* **18**, 579 (2016).
8. D. Finley, *Annu. Rev. Biochem.* **78**, 477 (2009).
9. A. L. Goldberg, *J. Cell Biol.* **199**, 583 (2012).
10. M. D. Ehlers, *Nat. Neurosci.* **6**, 231 (2003).
11. B. Bingol, E. M. Schuman, *Nature* **441**, 1144 (2006).
12. M. Schmidt, D. Finley, *Biochim. Biophys. Acta Mol. Cell Res.* **1843**, 13 (2014).
13. R. Fonseca *et al.*, *Neuron* **52**, 239 (2006).
14. K. V. Ramachandran, S. S. Margolis, *Nat. Struct. Mol. Biol.* **24**, 419 (2017).
15. A. J. George *et al.*, *Front. Genet.* **9**, 29 (2018).

10.1126/science.adi1166

MEDICINE

Emerging frontiers in regenerative medicine

Bridging knowledge gaps could enable regenerative therapy

By **Kara L. McKinley**[1,2], **Michael T. Longaker**[3,4], **Shruti Naik**[5,6,7,8]

Nearly every human malady, be it injury, infection, chronic disease, or degenerative disease, damages tissues (*1*). Moreover, 45% of all deaths can be traced to inflammation- and fibrosis-related regenerative failures (*1*). Restoring health after damage requires the answer to a key question: How can human tissues be coaxed to regenerate? Identifying instructive cues that direct refractory tissues down a regenerative path remains a critical yet elusive goal. Nonetheless, approaches to target roadblocks that impede regeneration, including insufficient and/or functionally inadequate progenitor cells, fibrosis, and chronic inflammation, are continuing to progress from bench to bedside. Pivotal advances have been made to overcome these hurdles using cell therapy, in vivo reprogramming, synthetic biology, and antifibrotic and anti-inflammatory therapies, but many challenges remain and knowledge gaps must be addressed to make regeneration a mainstay of modern medicine (*1*).

The most conspicuous requirement for regenerative therapies is to replace the components of tissues that were lost or compromised by disease. Invigorating endogenous stem cells is an appealing strategy, but, to date, the greatest benefits have emerged from cell therapies. Adult stem cell–based regenerative therapies have shown clinical benefit to treat hematological malignancies, burn wounds, and ocular degeneration. These therapeutic modalities also lend themselves to gene editing to correct monogenic perturbations. For instance, a young boy suffering from junctional epidermolysis bullosa, a lethal skin disease caused by mutations in the laminin-332 gene, was treated with an autologous (patient's own) skin transplant (*2*). A millimeter-sized sample of the boy's skin was collected and transduced with a retroviral vector expressing the wild-type laminin 332 cDNA; the tissue was then expanded ex vivo before being grafted to restore 80% of the body surface area.

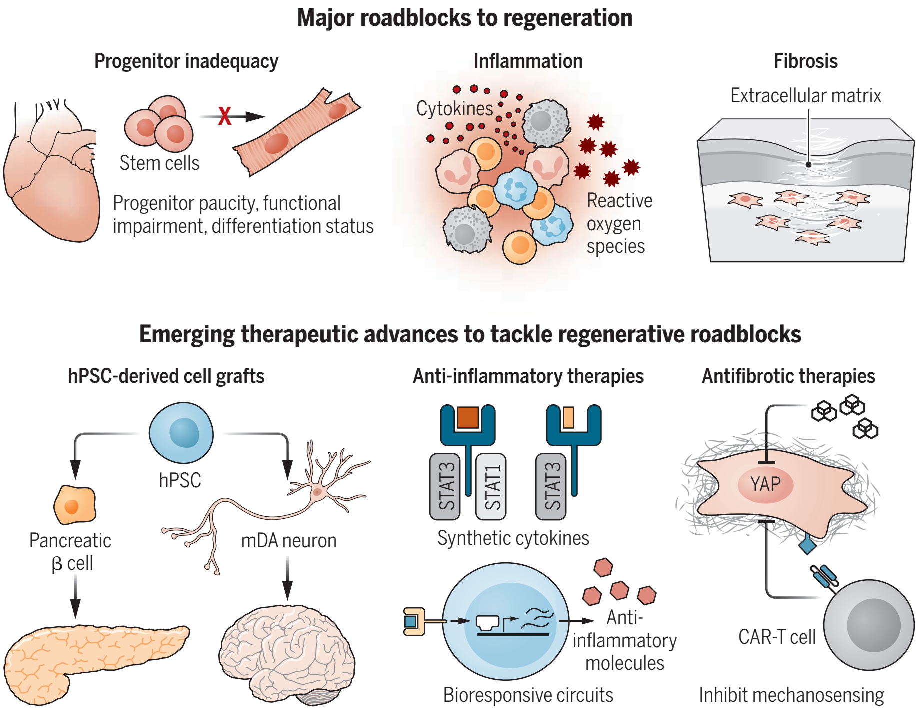
Human pluripotent stem cell (hPSC)–based therapies have also shown promise and have entered clinical trials in the United States for type 1 diabetes (T1D; clinical trial NCT04786262), Parkinson's disease (PD; NCT02452723 and NCT03119636), and age-related macular degeneration (AMD; NCT04339764) (*3*). These three diseases are particularly amenable to stem cell–based therapies because they are associated with deficiency of a defined cell type: insulin producing pancreatic islets in T1D, midbrain dopaminergic (mDA) neurons in PD, and retinal pigment epithelial cells in AMD. Moreover, transplantation sites are surgically accessible. Lab-generated hPSCs are being assessed for safety and efficacy in reducing disease symptoms. A case report of a PD patient who received midbrain transplant of autologous hPSC-derived mDA neurons showed stable grafting without immune reaction and exhibited improved motor function 2 years after implantation (*4*).

Despite these early glimpses of success, cell therapies are hampered by many biological and technical hurdles. Autologous hPSC-based therapies derived from induced pluripotent stem cells (iPSCs) avoid immune aggravation, but this cost- and labor-intensive strategy requires safety testing for each use. A major concern with nonautologous cell therapy is immune rejection of the graft. Thus, "off-the-shelf" allogeneic therapies must be coupled to strategies that allow them to avoid rejection, such as systemic immunosuppression in T1D hPSC therapy (NCT04786262). An alternative strategy is to encapsulate the graft within a selectively permeable barrier to allow for diffusion of nutrients and the therapeutic molecules (e.g., insulin) while excluding immune components that would reject the engrafted cells. To avoid fibrotic foreign

[1]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [2]Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA. [3]Division of Plastic and Reconstructive Surgery, Department of Surgery, Standford University School of Medicine, Stanford, CA, USA. [4]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. [5]Department of Pathology, New York University Langone Health, New York, NY, USA. [6]Department of Medicine, New York University Langone Health, New York, NY, USA. [7]Ronald O. Perelman Department of Dermatology, New York University Langone Health, New York, NY, USA. [8]Perlmutter Cancer Center, New York University Langone Health, New York, NY, USA. Email: kara_mckinley@harvard.edu; longaker@stanford.edu; shruti.naik@nyulangone.org

Therapeutic frontiers fueling advances in regenerative medicine

The three major roadblocks to regeneration are functional inadequacy of progenitors, chronic inflammation, and fibrosis. Ongoing clinical trials are testing the efficacy of hPSC-derived β cells and mDA neurons to treat type 1 diabetes and Parkinson's disease, respectively. Synthetic cytokines and bioresponsive systems are being developed to limit inflammation. Inhibiting mechanosensing pathways in fibroblasts and fibroblast antigen–specific CAR-T cells are being used to limit fibrosis.



CAR, chimeric antigen receptor; hPSC, human pluripotent stem cell; mDA, midbrain dopaminergic; STAT, signal transducer and activator of transcription; YAP, Yes-associated protein.

body responses, recent strategies for chemically modified alginate encapsulation showed promise in a nonhuman primate model. However, a phase 1 clinical trial that used a similar approach to deliver cells expressing factor VIII for the treatment of hemophilia A (NCT04541628) was halted after a patient experienced a serious adverse immune event and, subsequently, the implanted capsules were found to have fibrosed. Nonetheless, encapsulation strategies are moving forward in other contexts, including with recent US Food and Drug Administration (FDA) clearance for human trials of encapsulated stem cell–derived pancreatic islets (NCT05791201). In addition, ongoing work is investigating other strategies to reduce immune rejection, including genetic engineering of grafted cells to evade immune recognition or to secrete immune modulatory cytokines that induce localized immune tolerance.

Producing sufficient numbers of cells for engraftment has been successful in the skin (*2*) but poses a major challenge for regenerating tissue in other organs, particularly if only a small proportion of cells survive after transplant. For example, the survival rates of hPSC-derived mDA neurons is ~0.5 to 10%, and efforts are ongoing to identify the underlying mechanisms of graft loss. In addition, generating appropriate grafts when complex cell mixtures are required, such as in the lung to enable gas exchange, remains a challenge. Preclinical studies that leverage tissue self-assembly and organoid formation to transfer complex tissue units are underway.

Tissue overgrowth is a major concern of stem cell therapies because transferred cells can give rise to teratomas or other tumors. Growth-stop signals, particularly in the case of cell therapy, are not well understood, and molecular "off switches" will be vital to incorporate when deploying regenerative therapies. The Hippo signaling pathway was long considered a master developmental regulator of organ size because inactivating mutations in this pathway produced overgrowth phenotypes. Yet recent studies indicate that rather than engaging normal programs of development, Hippo pathway activation induces ectopic expression of aberrant growth pathways in *Drosophila melanogaster* (*5*). Whether developmental pathways are reengaged to rebuild adult tissues or adult regeneration requires a different set of regulators, as suggested by the study of Hippo signaling, is unclear. Thus, it will be necessary to carefully define microenvironmental and cell-intrinsic growth-restriction signals in age-appropriate and organ-specific models to restrain the growth of transplanted cells.

Instructive microenvironmental signals are often necessary for optimal cellular function. The success of stem cell therapies, which may be delivered intravenously, therefore depends on migration and integration into appropriate tissue niches. Cell or tissue grafts must also connect with the host's nervous and vascular machinery, and the signals involved remain poorly understood (*6*). An alternative approach to overcome the challenge of directing and integrating grafts in hard-to-reach internal organs is to repurpose cells that are present at the damage site by reprogramming them in situ with specific transcription factors (*6*). Reprogramming has been a particularly attractive strategy for the adult heart, which, unlike the embryonic heart, lacks bona fide progenitors. A subset of in vivo cellular reprogramming efforts are aimed at converting fibroblasts into cardiomyocytes to maintain cardiac function (*6*). Alternatively, transient expression of pluripotent transcription factors in adult cardiomyocytes induced proliferation, resulting in improved outcomes in adult mice after myocardial infarction (*7*). However, prolonged expression of these factors generated tumors. To overcome this substantial safety challenge, efforts to use enhancer elements first identified in highly regenerative zebrafish models to express reprogramming factors only in injury conditions are underway (*8*). Knowledge of which cells are reprogrammed, if and how they survive, what they become, and precisely how they contribute to clinical outcomes will be crucial to forecast treatment efficacy.

In addition to replacing lost tissue with cell therapies and in vivo reprogramming, many injurious and complex disease states evoke inflammatory responses that must also be suppressed (see the figure). The state of the recipient tissue, often diseased, remains a challenging facet of applying cell therapies. Cells or tissues may fail to engraft in inhospitable environments, or the underlying pathology that destroyed the original tissue may also damage the graft (*9*). For example, inflammatory bowel diseases (IBDs) are driven by aberrant immune responses, defective epithelial repair, loss of normal intestinal microbial communities, and translocation of bacteria into underlying tissue. Although supplying intestinal epithelial cells to the damaged intestines of an IBD patient may serve as a temporary patch, without mitigating inflammation or translocating microbes, these approaches face substantial hurdles to achieving long-term efficacy.

New approaches to modulate inflamma-

GRAPHIC: K. HOLOSKI/*SCIENCE*

tion include engineering stem cells with bioresponsive gene circuits that can sense inflammatory factors such as cytokines or reactive oxygen species and, in turn, induce the production of anti-inflammatory factors, allowing endogenous progenitors or transplanted cells to repair damage (*9*). Indeed, the beneficial effects of some stem cell therapies have been traced to immune-modulatory effects of the transplant, rather than the tissue-generating properties of the stem cells themselves. In a mouse model of cardiac ischemic injury, transplanting stem cells augmented cardiac function, not by production of cardiomyocytes but by activating macrophages that limit extracellular matrix (ECM) deposition and modulate the mechanical properties of an injured area to rejuvenate the heart (*10*).

Tempering the immune system presents a formidable challenge because the same factor can promote repair or inflict damage through inflammatory effects, depending on context (*11*). For instance, early in wounding, interleukin-17 (IL-17) enables hypoxia adaptation of damaged epithelium, but persistent IL-17 signaling potentiates pathology by recruiting damage-causing neutrophils (*11*). Synthetic biology provides opportunities to divorce the damage-causing effects of inflammation from those involved in repair. IL-22 induces the expression of proregenerative signal transducer and activator of transcription 3 (STAT3) and pro-inflammatory STAT1 transcription factors. Engineering IL-22 with altered receptor binding to only induce regenerative STAT3 boosted intestinal stem cell proliferation in a mouse model of radiation injury, without driving STAT1-mediated inflammatory outcomes (*11*).

A critical hurdle to regeneration is fibrosis, which rapidly plugs damaged tissue by haphazardly depositing ECM. Fibrosis profoundly compromises tissue mechanics and cellular interactions and physically obstructs organ function (*1*, *12*). Antiscarring therapies have been notoriously hard to achieve because profibrotic factors such as transforming growth factor–β (TGFβ) also have important functions in maintaining health. Alternatively, modulating mechanical signaling by inhibiting fibroblast Yes-associated protein (YAP), a mechanosensory transcription coregulator, prevented scarring during skin repair in mice (*12*). Notably, opposing fibrosis in this manner was sufficient to restore the skin's architecture and tensile strength. Similarly, using engineered chimeric antigen receptor (CAR) T cells to target conserved antigens on ECM-generating cardiac fibroblasts reduced fibrosis and revived heart function after ischemic injury (*13*). These studies indicate that adult organs may still possess molecular roadmaps to activate regenerative responses. A comparison of fibroblasts from regenerating reindeer antler skin and scarring back skin uncovered that inflammatory priming distinguishes the profibrotic state (*14*). Thus, precisely targeting inflammation may also ameliorate fibrosis and unlock latent regenerative capacity (*1*).

What was once considered the future of medicine is now becoming reality. But there is no magic pill for regeneration (yet). In addition to scientific and technological innovation, there are also practical considerations of cost and production. Innovations in regenerative therapies for complex diseases or damage involving multiple cell types have been hampered by the lack of appropriate preclinical models and a paucity of fundamental information on instructive signals to build tissues. Accordingly, efforts to systematically chart tissue repair over time, in different model systems, and after different types of damage are now underway (*14*). Rather than limiting therapies to the rules of mammalian physiology, radical strategies from nonvertebrate species and even the plant kingdom are surfacing. For example, nanosized plant photosynthetic systems that augment chondrocyte anabolism could limit cartilage degradation and osteoarthritis in mice (*15*). Finally, achieving regeneration in humans will require a rapid transition from rodent models to clinically relevant large animal and human studies. Ascending the summit of human regeneration demands an interdisciplinary effort that brings together biologists, biomedical engineers, and clinicians. The view from the top will reveal a transformed medical landscape that is able to seamlessly rejuvenate organs, ultimately extending human life span and health span. ■

REFERENCES AND NOTES

1. N. C. Henderson *et al.*, *Nature* **587**, 555 (2020).
2. T. Hirsch *et al.*, *Nature* **551**, 327 (2017).
3. T. W. Kim *et al.*, *Front. Cell Dev. Biol*. **8**, 729 (2020).
4. J. S. Schweitzer *et al.*, *N. Engl. J. Med*. **382**, 1926 (2020).
5. W. Kowalczyk *et al.*, *Science* **378**, eabg3679 (2022).
6. D. Srivastava, N. DeWitt, *Cell* **166**, 1386 (2016).
7. Y. Chen *et al.*, *Science* **373**, 1537 (2021).
8. R. Yan *et al.*, *Cell Stem Cell* **30**, 96 (2023).
9. F. Guilak *et al.*, *J. Orthop. Res*. **37**, 1287 (2019).
10. R. J. Vagnozzi *et al.*, *Nature* **577**, 405 (2020).
11. L. Guenin-Mace *et al.*, *Annu. Rev. Immunol.* **41**, 207 (2023).
12. H. E. Talbott *et al.*, *Cell Stem Cell* **29**, 1161 (2022).
13. J. G. Rurik *et al.*, *Science* **375**, 91 (2022).
14. S. Sinha *et al.*, *Cell* **185**, 4717 (2022).
15. P. Chen *et al.*, *Nature* **612**, 546 (2022).

ACKNOWLEDGMENTS

The authors thank the participants of the Janelia Research Campus "R3 – Replace, Repair, Regenerate" workshop (2021) for inspiring this perspective. S.N. is a New York Stem Cell Foundation Robertson Stem Cell investigator and is funded by National Institutes of Health (NIH) grants 1DP2AR079173-01 and R01-AI168462. K.L.M. is a New York Stem Cell Foundation Robertson Stem Cell investigator and is funded by NIH grants DP2HD111708 and R00HD101021. M.T.L. is supported by NIH grants R01-GM116892, R01- GM136659, and U24-DE02946. S.N. is on the scientific advisory board of Seed, Inc.; is a consultant for BiomX; and receives research funding from Takeda Pharmaceuticals.

10.1126/science.add6492

ECOLOGY

Past microbial stress benefits tree resilience

Soil microbiota from stressful environments provide an avenue for climate resilience

By **Michelle E. Afkhami**

By pushing environments to new extremes and exposing organisms to unprecedented levels of stress, anthropogenic changes are threatening biodiversity and ecosystem services. The substantial diversity and long evolutionary history of microorganisms provide a well of biological innovation that has the potential to relieve stress and increase ecosystem resilience (*1*). On page 835 of this issue, Allsup *et al.* (*2*) report that soil microbes can relieve climatic stress and enhance tree survival when the microbes have previous experience with that stress (drought or excess heat or cold). They also show that inoculated microbes, including beneficial mycorrhizal fungi, were still detectable in tree roots 3 years after planting in nature. These results suggest that management of soil microbiota, especially during restorations, could provide a valuable strategy for increasing forest resilience to climate change.

Microbial communities are fundamental to healthy, functioning environments around the globe. Soil microbiota underpin ecosystem services including nutrient cycling, decomposition, and carbon sequestration directly and through interactions with plants—the primary producers (autotrophs) that fuel food webs (*3*). Plants host diverse assemblages of fungi and prokaryotes (bacteria and archaea) that live on and inside roots, leaves, stems, and flowers. They can help plants withstand drought, high salinity, extreme heat and cold, low nutrients, heavy-metal pollution, and other challenging conditions (*4*). For example, the hyphal networks of mycorrhizal fungi in soils can access water and nutrients beyond the rhizosphere (the plant's rooting zone) to exchange for photosynthetic carbon, increasing fitness of both the fungi and plant.

Department of Biology, University of Miami, Coral Gables, FL, USA. Email: michelle.afkhami@miami.edu

tion include engineering stem cells with bioresponsive gene circuits that can sense inflammatory factors such as cytokines or reactive oxygen species and, in turn, induce the production of anti-inflammatory factors, allowing endogenous progenitors or transplanted cells to repair damage (*9*). Indeed, the beneficial effects of some stem cell therapies have been traced to immune-modulatory effects of the transplant, rather than the tissue-generating properties of the stem cells themselves. In a mouse model of cardiac ischemic injury, transplanting stem cells augmented cardiac function, not by production of cardiomyocytes but by activating macrophages that limit extracellular matrix (ECM) deposition and modulate the mechanical properties of an injured area to rejuvenate the heart (*10*).

Tempering the immune system presents a formidable challenge because the same factor can promote repair or inflict damage through inflammatory effects, depending on context (*11*). For instance, early in wounding, interleukin-17 (IL-17) enables hypoxia adaptation of damaged epithelium, but persistent IL-17 signaling potentiates pathology by recruiting damage-causing neutrophils (*11*). Synthetic biology provides opportunities to divorce the damage-causing effects of inflammation from those involved in repair. IL-22 induces the expression of proregenerative signal transducer and activator of transcription 3 (STAT3) and pro-inflammatory STAT1 transcription factors. Engineering IL-22 with altered receptor binding to only induce regenerative STAT3 boosted intestinal stem cell proliferation in a mouse model of radiation injury, without driving STAT1-mediated inflammatory outcomes (*11*).

A critical hurdle to regeneration is fibrosis, which rapidly plugs damaged tissue by haphazardly depositing ECM. Fibrosis profoundly compromises tissue mechanics and cellular interactions and physically obstructs organ function (*1*, *12*). Antiscarring therapies have been notoriously hard to achieve because profibrotic factors such as transforming growth factor–β (TGFβ) also have important functions in maintaining health. Alternatively, modulating mechanical signaling by inhibiting fibroblast Yes-associated protein (YAP), a mechanosensory transcription coregulator, prevented scarring during skin repair in mice (*12*). Notably, opposing fibrosis in this manner was sufficient to restore the skin's architecture and tensile strength. Similarly, using engineered chimeric antigen receptor (CAR) T cells to target conserved antigens on ECM-generating cardiac fibroblasts reduced fibrosis and revived heart function after ischemic injury (*13*). These studies indicate that adult organs may still possess molecular roadmaps to activate regenerative responses. A comparison of fibroblasts from regenerating reindeer antler skin and scarring back skin uncovered that inflammatory priming distinguishes the profibrotic state (*14*). Thus, precisely targeting inflammation may also ameliorate fibrosis and unlock latent regenerative capacity (*1*).

What was once considered the future of medicine is now becoming reality. But there is no magic pill for regeneration (yet). In addition to scientific and technological innovation, there are also practical considerations of cost and production. Innovations in regenerative therapies for complex diseases or damage involving multiple cell types have been hampered by the lack of appropriate preclinical models and a paucity of fundamental information on instructive signals to build tissues. Accordingly, efforts to systematically chart tissue repair over time, in different model systems, and after different types of damage are now underway (*14*). Rather than limiting therapies to the rules of mammalian physiology, radical strategies from nonvertebrate species and even the plant kingdom are surfacing. For example, nanosized plant photosynthetic systems that augment chondrocyte anabolism could limit cartilage degradation and osteoarthritis in mice (*15*). Finally, achieving regeneration in humans will require a rapid transition from rodent models to clinically relevant large animal and human studies. Ascending the summit of human regeneration demands an interdisciplinary effort that brings together biologists, biomedical engineers, and clinicians. The view from the top will reveal a transformed medical landscape that is able to seamlessly rejuvenate organs, ultimately extending human life span and health span. ■

REFERENCES AND NOTES

1. N. C. Henderson *et al*., *Nature* **587**, 555 (2020).
2. T. Hirsch *et al*., *Nature* **551**, 327 (2017).
3. T. W. Kim *et al*., *Front. Cell Dev. Biol*. **8**, 729 (2020).
4. J. S. Schweitzer *et al*., *N. Engl. J. Med*. **382**, 1926 (2020).
5. W. Kowalczyk *et al*., *Science* **378**, eabg3679 (2022).
6. D. Srivastava, N. DeWitt, *Cell* **166**, 1386 (2016).
7. Y. Chen *et al*., *Science* **373**, 1537 (2021).
8. R. Yan *et al*., *Cell Stem Cell* **30**, 96 (2023).
9. F. Guilak *et al*., *J. Orthop. Res*. **37**, 1287 (2019).
10. R. J. Vagnozzi *et al*., *Nature* **577**, 405 (2020).
11. L. Guenin-Mace *et al*., *Annu. Rev. Immunol*. **41**, 207 (2023).
12. H. E. Talbott *et al*., *Cell Stem Cell* **29**, 1161 (2022).
13. J. G. Rurik *et al*., *Science* **375**, 91 (2022).
14. S. Sinha *et al*., *Cell* **185**, 4717 (2022).
15. P. Chen *et al*., *Nature* **612**, 546 (2022).

ACKNOWLEDGMENTS

The authors thank the participants of the Janelia Research Campus "R3 – Replace, Repair, Regenerate" workshop (2021) for inspiring this perspective. S.N. is a New York Stem Cell Foundation Robertson Stem Cell investigator and is funded by National Institutes of Health (NIH) grants 1DP2AR079173-01 and R01-AI168462. K.L.M. is a New York Stem Cell Foundation Robertson Stem Cell investigator and is funded by NIH grants DP2HD111708 and R00HD101021. M.T.L. is supported by NIH grants R01-GM116892, R01- GM136659, and U24-DE02946. S.N. is on the scientific advisory board of Seed, Inc.; is a consultant for BiomX; and receives research funding from Takeda Pharmaceuticals.

10.1126/science.add6492

ECOLOGY

Past microbial stress benefits tree resilience

Soil microbiota from stressful environments provide an avenue for climate resilience

By **Michelle E. Afkhami**

By pushing environments to new extremes and exposing organisms to unprecedented levels of stress, anthropogenic changes are threatening biodiversity and ecosystem services. The substantial diversity and long evolutionary history of microorganisms provide a well of biological innovation that has the potential to relieve stress and increase ecosystem resilience (*1*). On page 835 of this issue, Allsup *et al.* (*2*) report that soil microbes can relieve climatic stress and enhance tree survival when the microbes have previous experience with that stress (drought or excess heat or cold). They also show that inoculated microbes, including beneficial mycorrhizal fungi, were still detectable in tree roots 3 years after planting in nature. These results suggest that management of soil microbiota, especially during restorations, could provide a valuable strategy for increasing forest resilience to climate change.

Microbial communities are fundamental to healthy, functioning environments around the globe. Soil microbiota underpin ecosystem services including nutrient cycling, decomposition, and carbon sequestration directly and through interactions with plants—the primary producers (autotrophs) that fuel food webs (*3*). Plants host diverse assemblages of fungi and prokaryotes (bacteria and archaea) that live on and inside roots, leaves, stems, and flowers. They can help plants withstand drought, high salinity, extreme heat and cold, low nutrients, heavy-metal pollution, and other challenging conditions (*4*). For example, the hyphal networks of mycorrhizal fungi in soils can access water and nutrients beyond the rhizosphere (the plant's rooting zone) to exchange for photosynthetic carbon, increasing fitness of both the fungi and plant.

Department of Biology, University of Miami, Coral Gables, FL, USA. Email: michelle.afkhami@miami.edu

As communities face escalating and new stresses in the Anthropocene, plants must find ways to persist in changing environments or track habitable conditions to avoid extinction. Microbiota can ameliorate stressors in situ or new stress encountered during range expansion (*5*).

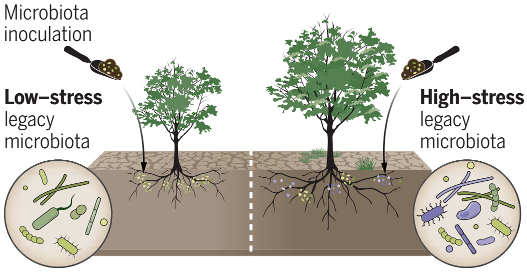
A legacy of stress can be good. Allsup *et al.* grew trees preinoculated with stress-experienced soil microbiotas from 12 sites across Wisconsin and Illinois in field and greenhouse common gardens for 3 years and evaluated tree survival, growth, and rhizosphere fungal composition. They demonstrated that soil microbiota experienced with climate stress—cold, hot, or dry conditions—are better able to promote tree survival under those specific conditions (e.g., warm-habitat microbiota ameliorate heat stress). If the stress legacy of microbiota enhancing their ability to promote tree health under that stress is a common phenomenon, this would provide a method for predicting when microbiota will facilitate climate resilience and for choosing the "correct" inoculates for use in restoration, management, and sustainable agriculture. Stress legacy effects have been supported in other systems, where only salinity-experienced microbiota ameliorated salt stress for mangroves (*6*, *7*). Further, the importance of stress legacy emphasizes that different habitats harbor distinct microbial communities that maximize benefits under those habitat conditions, contradicting the "everything is everywhere" adage. This highlights the need to consider how microbiota will migrate with climate change and whether active management of microbial communities is required for ecosystem health.

A key question moving forward is how stress primes the rhizosphere microbiota to ameliorate specific stresses in host plants (see the figure). One possibility is that stressful environments select for distinct groups of interacting fungi and prokaryotes that underpin functional benefits that are specific to that stressor, but which are absent when humans rapidly alter stress regimes. Another nonmutually exclusive explanation is that local adaptation of microbes to stressful conditions is crucial for microbes to maintain their own metabolic and physiological functions needed to produce and share benefits. Moreover, new interactions between indigenous and inoculum-introduced microbiota may allow better overall community functioning. For example, stress-tolerant microbes from inoculum could supply resources needed for indigenous microbiota health by replacing stress-sensitive species, or horizontal gene transfer could propagate stress-resistance adaptations within indigenous microbiomes. Future metagenomic and metatranscriptomic studies, ideally including single-cell sequencing, will be valuable for differentiating among these three possible mechanisms and for identifying how stress primes changes to function and dynamics in microbial communities.

How stress legacy of soil microbiota might increase tree resilience

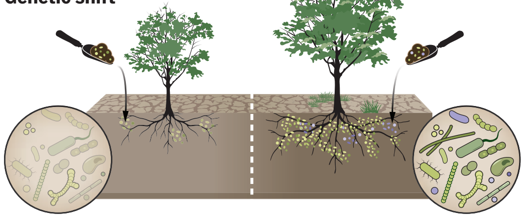
Inoculations with soil microorganisms can help trees withstand stressful conditions, such as drought, high salinity, soil contamination, and extreme temperatures, especially when they have previous experience with that stress (high stress legacy). Several possible mechanisms might mediate these beneficial effects under high stress such as drought.

Compositional shift



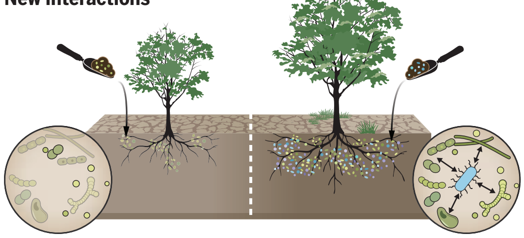
Stress legacy may select for microbiota compositional changes favoring microbes that provide functional benefits specific to that stressor.

Genetic shift



Stress legacy may promote genetic shifts that make microbiota better equipped to fully function under high stress.

New interactions



Interactions among indigenous and inoculum-introduced microbes may facilitate greater community-wide function.

GRAPHIC: N. CARY/*SCIENCE*

It is notable that Allsup *et al.* found that inoculated microbes were still detectable in experimental trees' roots after 3 years. This is both a point of excitement and caution for restoration. That properties of inoculated microbial communities and their effects on host plants were retained across years points to preinoculation of plants during revegetation efforts as a feasible avenue for restoring microbial communities and function (*6*). However, it also raises concerns because unplanned microbial introductions from nursery plants are likely common, and even carefully considered inoculations of seedlings or soils may lead to unintended consequences. For example, the longevity of microbes in plantings could facilitate spread of new species that displace indigenous microbial diversity, and horizontal gene transfer among introduced and indigenous microbes could rapidly reshape genetic variation in local communities.

Interestingly, in their northern sites, which simulated the expanding forest range edge, Allsup *et al.* found that no single microbiota optimized survival of trees experiencing both winter cold and summer drought. By contrast, trees at this site that only experienced cold stress (i.e., no drought) benefited from cold-experienced microbial inoculates. This suggests that multidimensional stress can disrupt the positive effect of microbial stress legacy. Similarly, recent work on soil microbiota showed that increasing the number of anthropogenic stressors disrupts soil microbial multifunctionality and can generate synergistic negative effects that cannot be predicted from individual stressor effects (*8*, *9*). Although natural ecosystems are inherently stressful across multiple dimensions, the introduction of new stressors or rapid intensification of anthropogenic stressors may contribute to disrupted microbial services by generating particularly strong conflicting selection on microbial functions, dismantling the fail-safe of functional redundancy within microbial communities and/or causing the loss of influential microbes. For example, "keystone microbes," which structure soil communities and provide important ecosystem functions (*10*), may be especially stress-sensitive (*11*). The finding that stress-experienced microbiomes can ameliorate climate stress raises hope for ecosystem resilience, but a comprehensive gene-to-ecosystems understanding of microbial roles in climate change resilience is needed before active management of soil microbial communities can be undertaken. ■

REFERENCES AND NOTES

1. C. Wagg *et al.*, *Nat. Commun.* **10**, 4841 (2019).
2. C. Allsup *et al.*, *Science* **380**, 835 (2023).
3. T. W. Crowther *et al.*, *Science* **365**, eaav0550 (2019).
4. S. Porter *et al.*, *Funct. Ecol.* **34**, 2075 (2020).
5. M. E. Afkhami *et al.*, *Ecol. Lett.* **17**, 1265 (2014).
6. J. Valliere *et al.*, *Ecol. Solut. Evid.* **1**, e12027 (2020).
7. S. C. Subedi *et al.*, *Ecology* **103**, e3679 (2022).
8. M. C. Rillig *et al.*, *Science* **366**, 886 (2019).
9. M. Rillig *et al.*, *Nat. Clim. Change* (2023).
10. M. T. Agler *et al.*, *PLOS Biol.* **14**, e1002352 (2016).
11. D. J. Hernandez *et al.*, *bioRxiv* 2023.03.17.533171 (2023).

10.1126/science.adi1594

MATERIALS SCIENCE

Plainly fixing crystal lattices

A thermoelectric alloy achieves high performance in electronic cooling

By In Chung[1,2]

Thermoelectrics is a green-energy technology that can use charge carrier transport within a material to achieve mutual conversion between thermal and electrical energy, consequently enabling both power generation and electronic cooling (*1*). It is relevant to various advanced technologies such as wireless communications, microelectronics, and medicine because thermoelectric (TE) devices have high mechanical reliability, can rapidly respond, and can vary in size and structure, depending on the application. On page 841 of this issue, Liu *et al.* (*2*) describe how tin selenide (SnSe) crystals can be improved as a TE material at room temperature by innovatively plaining the crystal lattice. The ground-breaking strategy produces a game-changer in seriously underdeveloped economical, eco-friendly, and noiseless cooling markets.

The cooling capacity of a thermoelectric cooler (TEC) is primarily determined by two parameters: the coefficient of performance (COP) and the maximum cooling temperature difference (ΔT_{max}). Both are closely related to the dimensionless figure of merit (*ZT*) of the cooling material near room temperature. *ZT* directly determines the conversion efficiency of TE devices and is defined as $S^2\sigma T/\kappa_{tot}$ (where S, σ, T, and κ_{tot} represent Seebeck coefficient, electrical conductivity, absolute temperature, and total thermal conductivity, respectively) (*3*). Because of the strong coupling relationships between these parameters in the competing electron and phonon transports within a TE material, it is impossible to boost one desirable parameter without adversely affecting the others. Accordingly, decoupling electron and phonon is essential to enhance the performance and cooling capacity of a TEC (*4*). However, performance optimization has remained a challenge.

To date, bismuth telluride (Bi_2Te_3)–based alloys are the only commercial TE cooling system and exhibit a *ZT* that is greater than unity near ambient temperature. However, their broad applications have been restricted by substandard performances, the sparse reserves of Te, and poor processability. Recently, materials such as Mg-based alloys (*5, 6*) and silver antimony telluride ($AgSbTe_2$) (*7*) showed room-temperature TE properties and cooling potential comparable to those of the Bi_2Te_3-based alloys. However, the highly reactive Mg as well as expensive Ag and Te elements make them inferior to Bi_2Te_3-based alloys. In this regard, SnSe is viewed as an important candidate to replace Bi_2Te_3-based alloys given the abundance of its constituents in Earth's crust. This makes SnSe a low-cost material with exceptional TE performances over a wide range of temperatures (*8–10*). Indeed, SnSe crystals have shown great promise as a commercial TE (*11, 12*). The outstanding question has been how to improve its TE performance at room temperature.

"The milestone lattice plainification strategy could enable a paradigm shift in the development of thermoelectric materials."

Liu *et al.* focused on manipulating the intrinsic defects in SnSe materials, especially Sn vacancies in these crystals, and provide an innovative solution to improve an otherwise highly imperfect crystal lattice. SnSe crystals have natural vacancies at the Sn position. These vacancies cause charge carriers that move through the crystal to scatter, which decreases the carrier mobility of SnSe. Curing these has been a sought-after goal to achieve high TE performance in SnSe materials (*8*). Liu *et al.* provide a new, highly effective solution in this study.

Introducing Cu atoms to the SnSe alloy fixes this defect by occupying these vacant positions, thus decreasing carrier scattering. This strategy achieved ultrahigh TE performance especially near room temperature, which is extremely rare in TE systems that normally operate in the intermediate temperature range (500 to 800 K). The obtained power factor of ~100 μW cm^{-1} K^{-2} at 300 K is exceptionally high, thereby enabling a room-temperature *ZT* of [illegible] and average *ZT* (ZT_{ave}) of ~2.2 in the 300 to 773 K range. The result is a high power generation efficiency of ~12.2% under a temperature difference between 298 and 598 K and a cooling ΔT_{max} of ~61.2 K near ambient temperature. This cutting-edge achievement was motivated by the "plainification" concept in which the properties and performances of alloys are enhanced by manipulating defects with less-alloyed compositions or even without introducing any foreign elements to materials, in contrast to traditional extrinsic alloying (*13*). In other words, properties of structural materials can be enhanced while their chemical compositions become simpler ("plainer").

Fixing up the crystal lattice corresponds to "plaining" it and is thus named "lattice plainification" by the authors. Notably, the Cu atom is also an effective dopant and promotes the multiband synglisis (momentum and energy multiband alignments) when entering the SnSe lattice, further optimizing the carrier transport and TE performance (*14*). Consequently, Cu facilitates better cooling performance of the material as well as lower internal resistance and power dissipation during the cooling operation. These achievements are essential for the commercialization of TECs (*15*).

The milestone lattice plainification strategy could enable a paradigm shift in the development of TE materials. It has been extremely difficult to modulate TE properties with respect to temperature. This strategy could be widely used to decouple electron–phonon transports in many TE materials, particularly those with massive intrinsic defects and with intrinsically low thermal conductivity. The resulting cheap TE cooling material is a breakthrough in opening new broad-based commercial applications. ■

REFERENCES AND NOTES

1. Q. Yan, M. G. Kanatzidis, *Nat. Mater.* **21**, 503 (2022).
2. D. Liu *et al.*, *Science* **380**, 841 (2023).
3. B. Jiang *et al.*, *Science* **377**, 208 (2022).
4. B. Qin, L.-D. Zhao, *Mater. Lab* **1**, 220004 (2022).
5. J. Mao *et al.*, *Science* **365**, 495 (2019).
6. J. Yang *et al.*, *Joule* **6**, 193 (2022).
7. S. Roychowdhury *et al.*, *Science* **371**, 722 (2021).
8. C. Zhou *et al.*, *Nat. Mater.* **20**, 1378 (2021).
9. Y. K. Lee, Z. Luo, S. P. Cho, M. G. Kanatzidis, I. Chung, *Joule* **3**, 719 (2019).
10. L.-D. Zhao *et al.*, *Nature* **508**, 373 (2014).
11. B. Qin *et al.*, *Nat. Commun.* **14**, 1366 (2023).
12. A. K. Munirathnappa, *Mater. Lab* **1**, 220056 (2022).
13. X. Li, K. Lu, *Science* **364**, 733 (2019).
14. B. Qin *et al.*, *Science* **373**, 556 (2021).
15. B. Qin, L.-D. Zhao, *Science* **378**, 832 (2022).

ACKNOWLEDGMENTS

The author acknowledges support from the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (no. NRF-2020R1A2C2011111) and the Institute for Basic Science (IBS-R009-D2).

10.1126/science.adi2174

[1]School of Chemical and Biological Engineering and Institute of Chemical Processes, Seoul National University, Seoul 08826, Republic of Korea. [2]Center for Correlated Electron Systems, Institute for Basic Science, Seoul 08826, Republic of Korea. Email: inchung@snu.ac.kr



INSIGHTS

POLICY FORUM

INTERNATIONAL LAW

Implementation, compliance, and pandemic legal obligations

Negotiations ought not focus on enforcement and sanctions

By **Mark Eccleston-Turner[1], Gian-Luca Burci[2], Jonathan Liberman[3], Sharifah Sekalala[4]**

Member states of the World Health Organization (WHO) are undertaking ambitious governance reforms to prevent, prepare for, and respond to pandemics by concurrently negotiating both a new international legal instrument (henceforth called "the Pandemic Treaty") and amendments to the International Health Regulations (IHR). We therefore find ourselves at a critical juncture in global health governance, with the opportunity to strengthen pandemic prevention, preparedness, and response through international law. One issue being given considerable prominence in each of these negotiations is how compliance by member states with their obligations can best be achieved. We argue that any efforts to ensure compliance with these instruments should be seen as part of broader efforts to ensure effective and equitable implementation, as opposed to being overly focused on formal compliance mechanisms and the possibility of punitive action in response to noncompliance.

Though there is considerable overlap between the IHR and the Pandemic Treaty, and the focus of their reform efforts, the push for a separate treaty has been based on the idea that the global response to a pandemic event, and considerations of equity during that event, are better addressed outside the IHR, which are overtly focused on preparedness and response. Compliance has dogged the IHR since their inception in 1969, notably during COVID-19, as widespread noncompliance saw numerous states acting in their own perceived self-interest, rendering WHO unable to coordinate an effective international response. This included failing to promptly share information with WHO and the international community and imposing disproportionate trade and travel restrictions against WHO advice. Moreover, many countries lacked the capacity to adequately control the pandemic—also an IHR compliance issue. Claims that the IHR have been ineffective during COVID-19 primarily owing to failures of compliance and enforcement (as opposed to, for example, reflecting gaps in the substantive provisions of the Regulations, and/or a lack of national capacities or support for implementation) have become commonplace (*1, 2*). Such calls appear to be finding traction in the current negotiations. Amendments to IHR have been proposed to both strengthen obligations and associated compliance mechanisms for preparedness and response, including through the creation of compliance assessment bodies. Additionally, the "zero draft" of the Pandemic Treaty proposes establishing "a universal peer review mechanism" and "cooperative procedures and institutional mechanisms to promote compliance…and also address cases

[1]Department of Global Health and Social Medicine, King's College London, London, UK. [2]Graduate Institute of International and Development Studies, Geneva, Switzerland. [3]Schools of Law and Population and Global Health, University of Melbourne, Melbourne, Australia. [4]Department of Law, University of Warwick, Coventry, UK. Email: mark.eccleston-turner@kcl.ac.uk

of non-compliance" (3). The exact formulation of such mechanisms is yet to be agreed. There have also been high-profile calls to strengthen accountability mechanisms to achieve greater compliance.

We challenge the empirical basis for some of these arguments and argue that the current focus on compliance in the literature, as well as the negotiations, often fails to reflect that the ultimate aim of agreeing new treaties is to implement them effectively. This means looking to what the Treaty intends to achieve as an overall package, rather than merely looking to the binary obligations contained within. This goes well beyond formal "compliance." For example, a focus on compliance can often become an overly legalistic search for, and concentration on, what an instrument specifically legally requires (as opposed to encourages or empowers), which might include an overemphasis on "shall" provisions, and a downplaying of the importance of "shall consider" and "should" provisions and of the treaty's overarching purpose and norms. By contrast, an implementation focus can often provide more space for identifying how best practice implementation of all commitments, however expressed, can be supported and facilitated, including distinguishing between different kinds of commitments, and the reasons that different obligations might be breached, or at least not optimally implemented. A "compliance" focus may suit some types of obligations—for example, those where obligations are clear, and contraventions are likely to be deliberate, such as failures to notify known information—but may not suit others, particularly where the underlying issue might be a lack of capacity.

Also problematic with this view is that states clearly consider a much broader range of responses to be punitive in nature—not just formal sanctions. For example, some states appear to view it as punitive if an event in their territory is declared a "public health emergency of international concern" (PHEIC), akin to WHO issuing a vote of no confidence in the state's ability to adequately respond. Equally, though not a formal sanction, imposition of trade and travel restrictions following declaration of a PHEIC, or following the notification by a state of public health information evidencing a risk, can be highly punitive in its impacts (*4*). Such consequences may discourage both compliance and best practice implementation.

COMPLIANCE IN INTERNATIONAL LAW

It is argued by some (*5*), including the Independent Panel on Pandemic Preparedness and Response (*6*), that poor IHR compliance is largely due to the lack of effective compliance mechanisms within the Regulations (*7*) and that this could be rectified by the inclusion of such mechanisms (or even sanctions) in the Regulations and/or the Pandemic Treaty. Some have gone so far as to claim that "enforcement mechanisms appear to be the only treaty design choice that holds promise of maximizing the chances of achieving intended effects" and that treaties "that do not have enforcement mechanisms are unlikely to be worth their considerable effort" (*2*, p. 6). For instance, human rights treaties come under considerable criticism for making human rights outcomes worse, with a particular example of the UN Convention on the Rights of the Child, which was found to have harmful effects because it led to, among others, worsened human rights practices and increases in child labor (*2*). However, this seems to ignore that before the signing of the treaty, there were considerable disagreements about the concept of child labor, and that it feels so abhorrent today in large part owing to the socialization effects of the treaty, which increased the rates of reporting of behavior that is perceived to be contrary to the treaty. Additionally, even these scholars acknowledge that effectiveness of a treaty can be enhanced through socialization, and changing norms (*2*). This is not a linear process and is achieved through development of consensus and internalization of norms as opposed to enforcement (*8*). Additionally, scholars have argued that states may comply because of the fear of reputational damage if they derogate from agreed norms (*9*). Thus, compliance in this sense is not binary but a spectrum, which may necessitate ongoing development of the norms of treaties over time (*10*).

We therefore caution against absolutism in respect to the utility of formal compliance or enforcement mechanisms, because it is implementation that matters most, not formal compliance systems designed to punish errant behavior, and also, because "hard" compliance mechanisms go against the norm within international law, implying that it is unrealistic to expect states to agree to such mechanisms. Most treaties lack hard compliance mechanisms, and treaties that do contain them are often bilateral (e.g., arms control) or part of a broader regulatory framework (e.g., trade, climate change, and protection of the ozone layer). For the most part, compliance and enforcement mechanisms are confined to reporting, some form of review powers exercised by intergovernmental or expert bodies, and, rarely, formal dispute settlement mechanisms (such as for the World Trade Organization) (*2*). A notable trend in environmental law is in favor of compliance mechanisms based on dialogue, identification of problems, and mutual support, which are considered a useful approach to support implementation and compliance across a range of obligations. Even these forms of "soft" compliance can act as a deterrence against noncompliance, whereby states are concerned about public disclosures of their failures to comply with obligations.

In short, approaches to compliance need not be strict to be meaningful. Chayes and Chayes noted that standards of strict, absolute compliance are often ill-suited to international law, and that it is more realistic to think of an "acceptable" overall level of compliance "in the light of the interests and concerns the treaty is designed to safeguard" (*11*, p. 176). They noted that "What is an acceptable level of compliance will shift according to the type of treaty, the context, the exact behaviour involved, and over time" (*11*, p. 198).

To this end, it has become something of a cliché to quote that "almost all nations observe almost all principles of international law and almost all of their obligations almost all of the time" (*12*, p. 47), but why noncompliance occurs is given insufficient emphasis in the literature, and has tended to be oversimplified or ignored in negotiations thus far. It is not always the case that noncompliance is nefarious in nature, or a calculated action taken by a state pursuing its own self-interest. Although such action is possible, particularly during an emergency event, it should not be the sole focus of efforts around implementation and compliance. On many occasions, states may deviate from treaty obligations because of a lack of resources—financial, knowledge, human, or structural—or genuine uncertainty about what the treaty requires, or knowledge gaps and coordination problems within the legal regime. This is the case for noncompliance with the prevention, detection, and response "Core Capacities" (*13*) requirements under the IHR and may also be the case for future resource-intensive obligations under the Pandemic Treaty.

ESSENTIAL CRITERIA

Negotiators should try to reach satisfactory trade-offs and balance between rights and obligations, and mechanisms for implementation support and for compliance. Negotiators ought to carefully consider the optimal role of formal compliance mechanisms, especially during times of crisis where attention and resources may be better directed elsewhere, and to recognize such mechanisms as one part of the overall toolbox that is designed to support and facilitate implementation. This would demand, in the first instance, considering why noncompliance might occur, including the role of both resource scarcity and knowledge gaps. A holistic approach to implementation could include features such as regular Conferences of the Parties (COPs), whereby state parties can network, share best practices, make commitments to each other outside of the formal renegotiation of an in-

strument, and adopt a range of normative instruments needed to support implementation. For example, in the case of WHO's Framework Convention on Tobacco Control, the COP has adopted several implementation guidelines that clarify and elaborate upon often generally expressed treaty obligations, drawing on the best available, and often evolving, evidence (*14*). Such precise and detailed—albeit softer—instruments, can generate higher levels of compliance (*15*). In multilateral environmental agreements, there is a particular emphasis on COPs being used for mutual learning and support, identification of specific and systemic problems, and linkage with technical and financial assistance to facilitate implementation and compliance.

Negotiators ought to be attentive to the inevitable costs of any compliance regime they create, especially for countries from the Global South. Formal compliance mechanisms that cover the wide range of obligations in the IHR, as well as those currently proposed for the Pandemic Treaty, would entail high costs—in addition to the costs of compliance itself, such as with IHR Core Capacities requirements. The "reporting fatigue" arising from multiple and increasing treaty obligations is not a new issue and is, for example, affecting credible compliance monitoring of UN human rights treaties.

Moreover, while negotiators are currently focused on obligations contained within the amended IHR and new Pandemic Treaty, and how to maximize compliance with these two instruments, it ought not be forgotten that other international law regimes that are also relevant to pandemic prevention, preparedness, and response—e.g., environment, climate change, animal health, trade, and intellectual property—are placing demands on states. Negotiators need to recognize that states will be "servicing" the cost and compliance requirements of such other regimes.

Finally, any holistic implementation regime should establish a governance framework that creates a sense of community including governments and nonstate actors, offers a space for discussion and review, and allows decisions on implementation and compliance to be taken in a dynamic manner, responding to the way in which circumstances and evidence, and implementation of the instruments themselves, evolve. Negotiators ought to avoid overly detailed compliance provisions within the text of the instrument itself, and instead use the instrument to either create the compliance body or entrust this authority to the COP (or the World Health Assembly in case of the IHR), leaving details of its composition and operation to be agreed later by these bodies.

Such an approach will provide states with a greater sense of flexibility and responsiveness in the compliance mechanism, informed by the early experiences of implementation, rather than seeking to establish a compliance regime for every obligation agreed to during the negotiations. This is particularly important, given the short timelines for both the IHR amendment and Treaty negotiations—with adoption of each currently planned for the May 2024 session of the World Health Assembly—and the sheer scale of potential obligations still on the negotiating table. This will also allow states to explore how they wish to ensure the overall effectiveness of the legal regime by using other mechanisms to facilitate both best practice implementation and legal compliance, such as using soft mechanisms not backed up by the prospect of punitive action. These can range from the very soft such as meetings of ministers to discuss progress on compliance, to more concrete mechanisms that can be undertaken either by the WHO or nongovernmental organizations, such as benchmarking indexes against commonly agreed indicators.

Additionally, providing financial and technical support for implementation, especially for countries in the Global South to strengthen national capacities, would potentially be a major incentive toward compliance. However, strong demands for "equity" in the current negotiations show that such support cannot be a form of "charity" and has to move away from a development assistance model. To this end some low- and middle-income countries are pushing for "common but differentiated responsibilities" being a key element of the Treaty, whereby high-income countries bear a greater responsibility to address the substantive goals of the Treaty than low-income countries. Finding common ground will be a major challenge for negotiators, but essential for the success of the emerging normative framework.

MORE HOLISTIC APPROACHES

Compliance has traditionally been one of the thorniest questions in international law, both theoretically and practically, and will remain so long after these reforms have reached their conclusions. How should we understand compliance with international obligations given the infrequency of enforcement mechanisms and the traditional dominance of rational choice theories anticipating noncompliance if a cost-benefit calculation points states in that direction? And from a practical perspective, what can be done to create the basis for better compliance and, more importantly, for effective and equitable implementation of treaties, in particular those of a regulatory nature where there is no immediate reciprocity of benefits? The short analysis in this article defies simplistic assumptions about one-size-fits-all approaches or the absolute necessity of highly formalized and potentially antagonistic mechanisms backed up by the underlying threat of sanctions. The evidence on such approaches is clear: States almost never agree to such compliance regimes being included within treaty regimes. Therefore, the international community needs to look to more holistic approaches to implementation and compliance, considering the overall legal, institutional and political "package" and whether and how it can generate trust and mutual reliance.

In the case of the Pandemic Treaty and the IHR amendments, the rather complex texts available as of the time of writing reveal tension between the priority for health security mechanisms (e.g., information and pathogen sharing and early notification of events) for countries of the Global North; and an equally urgent claim for equity and fairness (e.g., guaranteed access to vaccines, medicines, and other medical countermeasures; transfer of technology; financing) by countries of the Global South. If negotiators succeed in achieving a "package deal" that merges those two claims as cohesive and complementary components of the Pandemic Treaty and strengthened IHR, it will in our view offer the best chance of an effective, equitable, and satisfactorily implemented—and complied with—international regime on pandemic prevention, preparedness, and response. ■

REFERENCES AND NOTES

1. A. Spagnolo, *Global Jurist* **18**, 20170025 (2018).
2. S. J. Hoffman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2122854119 (2022).
3. WHO, "Zero draft of the WHO CA+ for the consideration of the Intergovernmental Negotiating Body at its fourth meeting;" https://apps.who.int/gb/inb/pdf_files/inb4/A_INB4_3-en.pdf.
4. M. Eccleston-Turner, C. Wenham, *Declaring a Public Health Emergency of International Concern: Between International Law and Politics* (Bristol Univ. Press, 2021).
5. L. O. Gostin, R. Katz, *Milbank Q.* **94**, 264 (2016).
6. The Independent Panel for Pandemic Preparedness and Response, "COVID-19: Make it the last pandemic" (2021); https://theindependentpanel.org/wp-content/uploads/2021/05/COVID-19-Make-it-the-Last-Pandemic_final.pdf.
7. M. Sohn, D. Ro, D. H. Koh, S. Lee, S. Y. Kim, *J. Glob. Health Sci.* **3**, e18 (2021).
8. A. T. Guzman, *Calif. Law Rev.* **90**, 1823 (2002).
9. M. Finnemore, K. Sikkink, *Int. Organ.* **52**, 887 (1998).
10. J. von Stein, "Compliance With International Law" in *Oxford Research Encyclopedia of International Studies* (Oxford Univ. Press, 2010); http://oxfordre.com/internationalstudies/view/10.1093/acrefore/9780190846626.001.0001/acrefore-9780190846626-e-55.
11. A. Chayes, A. H. Chayes, *Int. Organ.* **47**, 175 (1993).
12. L. Henkin, *How Nations Behave* (Columbia Univ. Press, ed. 2, 1979).
13. "Core capacities" is a term of art for the obligation of IHR states parties to develop and maintain the capacity in their health systems to detect, assess, and respond to disease outbreaks, as stipulated in Articles 5 and 13, and Annex 1, of the Regulations.
14. Secretariat to the WHO Framework Convention on Tobacco Control, "Guidelines, and policy options and recommendations for implementation of the WHO FCTC" (2022); https://fctc.who.int/who-fctc/overview/treaty-instruments.
15. K. W. Abbott, D. Snidal, *Int. Organ.* **54**, 421s (2000).

10.1126/science.adh2080



Aluminum ingots await use in a stockyard in Wuxi, Jiangsu province, China.

PHOTO: QILAI SHEN/BLOOMBERG VIA GETTY IMAGES

MINERAL SCIENCE

Ode to aluminum

An expansive history offers insights into the abundant metal's many facets

By **Katrin Daehn**

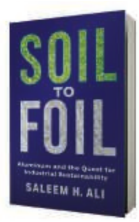
While most of us appreciate that minerals are essential to our economies and technologies, we tend to place them in a separate category from biological life. However, life and minerals are inextricably intertwined. This relationship is best revealed with an expansive examination of our connected fates that traverses history at various scales. With approachable storytelling, in *Soil to Foil*, environmental scientist Saleem Ali masterfully traces one such story—the story of aluminum—structuring the book around the mineral's natural occurrence and its roles in natural systems, its discovery and the scaling of its extraction, the products in which it appears and their uses, and possible strategies to establish circular use and reuse and to restore ecological damage caused by its extraction.

Mineral constraints are ultimately set by a substance's abundance and availability in Earth's crust, factors that are largely determined by cosmic events. Our planet's proximity to the Sun and its frequent collisions with meteors have favored the presence of metals with high melting points.

Aluminum, in particular, has a notable affinity for oxygen, so as elements redistributed during the Hadean eon, aluminum became the third most concentrated element in Earth's crust. Aluminum's affinity for oxygen resulted in its abundance but also long thwarted attempts to isolate elemental aluminum, which today is accomplished by energy-intensive smelting. Once extracted, aluminum forms an impervious, extremely thin oxide layer in air that is a key feature of its durability and perhaps its long-term sustainability.

Aluminum is a remarkable element that embodies several paradoxes—it is ubiquitous, yet rarely highly concentrated in the Earth's crust; it is infinitely recyclable in theory, but highly susceptible to impurities and degradation in practice; it is an ideal material for many household uses, but with still-debated biological consequences. It thus provides a rich context to wrestle with the difficult questions involved in transitioning toward "industrial sustainability." In the end, there are no easy answers, but we must keep engaging in the hard questions to establish practices for mineral extraction and use that are sustaining and life-supporting.

In the book's preface, Ali states that navigating between "need" and "greed" at the individual and collective levels is the great ecological challenge with regard to minerals. The boundaries that define this space are determined by resource constraints and human agency, the latter of which includes our technological and governance capability. Aluminum offers insight on each of these themes that may be useful to a range of other resources and in other contexts.



Soil to Foil: Aluminum and the Quest for Industrial Sustainability
Saleem H. Ali
Columbia University Press, 2023. 320 pp.

The vast power and wealth created by mineral markets have long been present in the undercurrent of geopolitics. Alum, for example—an aluminum-containing salt used to impart a prized purple color and fire-retardance to clothing during the 15th century—was of particular interest to Crusaders and partially motivated their invasions and conquests. The Catholic Church, we learn, went on to form an alum monopoly, and at its height, one-third of the revenue of the papacy was derived from the prized salt. A chance discovery that urine added to heated rocks produced ammonium alum finally provided a pathway for dissent against the papacy, transforming the fate of European history.

Aluminum production scaled up in the 20th century, just in time for the metal to become a major player in World Wars I and II, where it was used to create strong but lightweight weapons and aircraft. In the postwar era, a portfolio of new civilian products was invented and promoted with clever marketing. Vast swathes of land, ecosystems, and communities have since been altered to continuously increase aluminum production, which Ali illuminates with detail and care. However, as he also shows, aluminum sustains countless livelihoods in industries supported by this commodity supply chain. It has raised human technological capability and is now part of the fabric of modern life.

One of the most urgent tasks for humanity is to combat climate change, and Ali explores a variety of options that could aid in this effort, from increasing aluminum recycling, to restorative "phytomining" (a technique for harvesting metals from plants that accumulate them at high levels), to new models of mine ownership that could bring wealth to the communities in which they are situated.

As we stand on the precipice of electrification with renewable energy, it is intriguing to think how this supply of electricity may change the paradigm for other materials separations. Books like *Soil to Foil* remind us not to lose sight of the vital roles minerals have played and continue to play in our society. ■

10.1126/science.adh4905

The reviewer is at the MIT Climate & Sustainability Consortium, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. Email: daehn@mit.edu

CLIMATE CHANGE

Adding fuel to the fires

A wildfire in a Canadian oil town offers a case study in fossil fuels and climate change

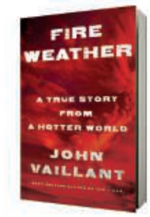
By **Sarah Boon**

On 1 May 2016, a human-caused wildfire ignited in the forest east of Fort McMurray, a northern Alberta "petro-city" of ~88,000 residents that is sustained by oil sands mining. Named the Horse River Fire, also called "The Beast," within days it had jumped the Athabasca River and made its first of three runs into the city. The fire burned 2300 square miles in all and was not declared under control until 5 July 2016. It was finally extinguished on 2 August 2017, 15 months after it started. In total, 2400 homes and other structures were destroyed, and thousands more were damaged. Astonishingly, no one was injured. All of the area's residents survived the chaotic evacuation, "the largest, most rapid single-day evacuation in the history of modern fire." It was the most expensive natural disaster in Canadian history.

In his latest book, *Fire Weather*, Canadian author John Vaillant describes the city's response to the fire as haphazard and unprepared. "It is hard to overstate how unconcerned was Fort McMurray's citizenry," he writes, wondering whether the slow response was related to residents' can-do, bootstrapping mentality. There was also the oil industry to consider, as "no investor or CEO was going to want to hear about a shutdown due to fire."

The book's immediacy and forward momentum come from Vaillant's extensive firsthand interviews with residents and firefighters who were on the ground the day the fire began and those who returned after the fire to survey the devastation. "There's my daughter, there's my wife, there's a fireball going over top of the truck and my wife's driving into it. That's intense heat," said one evacuee. As an official put it, "You go to a place where there was a house and what do you see on the ground? Nails. Piles and piles of nails." Vaillant also uses imagery to great effect, for example, describing the firefighting efforts as "[not] so much a fire alarm as it was the mobilization of a militia in the face of an invading army."

In addition to providing a play-by-play of how the fire evolved and its devastating aftermath, Vaillant probes the history of the development of the oil sands on which Fort McMurray is built and explains how the bitumen industry has thrived on Alberta's "low taxes, generous subsidies, and minimal oversight." He also addresses the history of climate change science, which oil companies were aware of as early as the 1950s and began working to suppress in the 1980s because of the unfavorable pall it cast on the oil industry. Many turned instead to disinformation campaigns that sought to sell the benefits of fossil fuels and minimize their drawbacks, he writes.



Fire Weather: A True Story from a Hotter World
John Vaillant
Knopf, 2023. 432 pp.

Vaillant explains that wildfire risk is calculated using a trifecta of factors: heat (air temperature), fuel ("duff"—dry needles, cones, and branches—and flammable trees), and oxygen. He returns to this theme of threes later in the book, when discussing how the oil industry exacerbates wildfire risk by burning fossil fuels, which leads to the emission of CO_2, which promotes climate change. The result is hotter and bigger forest fires. Fort McMurray burned at least in part because of the industry on which it relies, he argues.

Vaillant describes fossil fuel pollution as the ultimate example of the tragedy of the commons. It is also, he argues, a form of colonization, whose impacts are far-reaching, with outsized consequences for those in the Global South. Fort McMurray itself is built on a colonial model, he maintains, that "systematically commodifies natural resources and binds local people to the trading post system with company store–style debt."

But Vaillant sees cause for hope even as CO_2 emissions rise. Fossil fuel companies are suffering as large organizations divest their portfolios of oil and gas stocks. They are also finding that, "with the low-hanging fruit long gone, the cost of exploration and recovery is only growing."

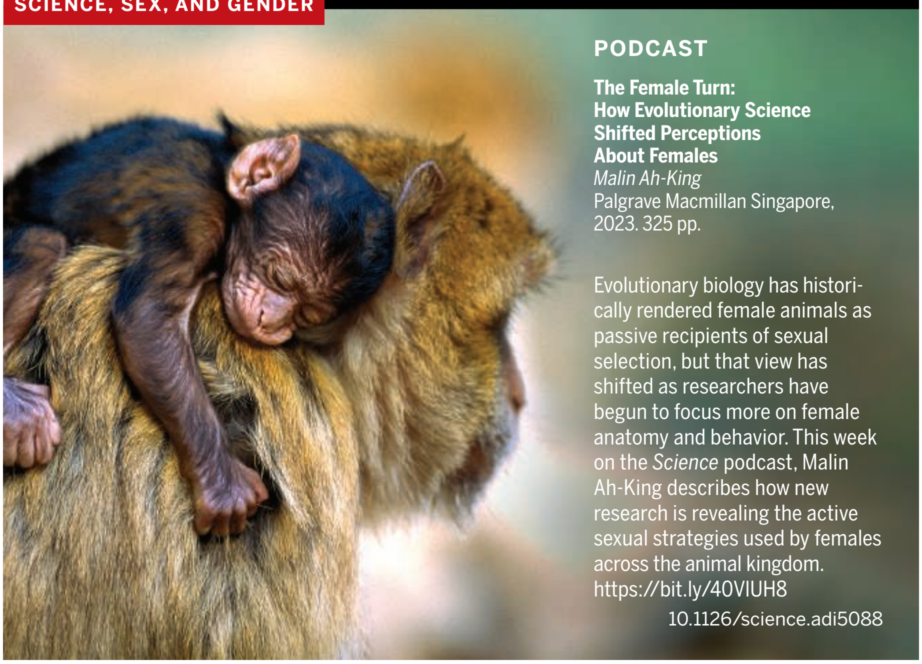
He also sees hope in Fort McMurray's recovery from the fire, even though ~20,000 residents never returned. He describes the reviresence of the burned city: "the forest around Fort Mac is slowly reviving…and the leveled neighborhoods are filling in with new homes, trees, and flowers."

Fire Weather is a gripping book that brings readers to the front lines of a major forest fire, while also exploring the intertwined history of oil and gas development and the study of climate change. Its lessons should not be soon forgotten. ■

10.1126/science.adh9238

The reviewer is a freelance science writer and editor and cofounder of *Science Borealis*. Email: snowhydro1@gmail.com

SCIENCE, SEX, AND GENDER



PODCAST

The Female Turn: How Evolutionary Science Shifted Perceptions About Females
Malin Ah-King
Palgrave Macmillan Singapore, 2023. 325 pp.

Evolutionary biology has historically rendered female animals as passive recipients of sexual selection, but that view has shifted as researchers have begun to focus more on female anatomy and behavior. This week on the *Science* podcast, Malin Ah-King describes how new research is revealing the active sexual strategies used by females across the animal kingdom. https://bit.ly/40VIUH8

10.1126/science.adi5088

CLIMATE CHANGE

Adding fuel to the fires

A wildfire in a Canadian oil town offers a case study in fossil fuels and climate change

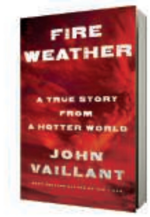
By **Sarah Boon**

On 1 May 2016, a human-caused wildfire ignited in the forest east of Fort McMurray, a northern Alberta "petro-city" of ~88,000 residents that is sustained by oil sands mining. Named the Horse River Fire, also called "The Beast," within days it had jumped the Athabasca River and made its first of three runs into the city. The fire burned 2300 square miles in all and was not declared under control until 5 July 2016. It was finally extinguished on 2 August 2017, 15 months after it started. In total, 2400 homes and other structures were destroyed, and thousands more were damaged. Astonishingly, no one was injured. All of the area's residents survived the chaotic evacuation, "the largest, most rapid single-day evacuation in the history of modern fire." It was the most expensive natural disaster in Canadian history.

In his latest book, *Fire Weather*, Canadian author John Vaillant describes the city's response to the fire as haphazard and unprepared. "It is hard to overstate how unconcerned was Fort McMurray's citizenry," he writes, wondering whether the slow response was related to residents' can-do, bootstrapping mentality. There was also the oil industry to consider, as "no investor or CEO was going to want to hear about a shutdown due to fire."

The book's immediacy and forward momentum come from Vaillant's extensive firsthand interviews with residents and firefighters who were on the ground the day the fire began and those who returned after the fire to survey the devastation. "There's my daughter, there's my wife, there's a fireball going over top of the truck and my wife's driving into it. That's intense heat," said one evacuee. As an official put it, "You go to a place where there was a house and what do you see on the ground? Nails. Piles and piles of nails." Vaillant also uses imagery to great effect, for example, describing the firefighting efforts as "[not] so much a fire alarm as it was the mobilization of a militia in the face of an invading army."

In addition to providing a play-by-play of how the fire evolved and its devastating aftermath, Vaillant probes the history of the development of the oil sands on which Fort McMurray is built and explains how the bitumen industry has thrived on Alberta's "low taxes, generous subsidies, and minimal oversight." He also addresses the history of climate change science, which oil companies were aware of as early as the 1950s and began working to suppress in the 1980s because of the unfavorable pall it cast on the oil industry. Many turned instead to disinformation campaigns that sought to sell the benefits of fossil fuels and minimize their drawbacks, he writes.



Fire Weather: A True Story from a Hotter World
John Vaillant
Knopf, 2023. 432 pp.

Vaillant explains that wildfire risk is calculated using a trifecta of factors: heat (air temperature), fuel ("duff"—dry needles, cones, and branches—and flammable trees), and oxygen. He returns to this theme of threes later in the book, when discussing how the oil industry exacerbates wildfire risk by burning fossil fuels, which leads to the emission of CO_2, which promotes climate change. The result is hotter and bigger forest fires. Fort McMurray burned at least in part because of the industry on which it relies, he argues.

Vaillant describes fossil fuel pollution as the ultimate example of the tragedy of the commons. It is also, he argues, a form of colonization, whose impacts are far-reaching, with outsized consequences for those in the Global South. Fort McMurray itself is built on a colonial model, he maintains, that "systematically commodifies natural resources and binds local people to the trading post system with company store–style debt."

But Vaillant sees cause for hope even as CO_2 emissions rise. Fossil fuel companies are suffering as large organizations divest their portfolios of oil and gas stocks. They are also finding that, "with the low-hanging fruit long gone, the cost of exploration and recovery is only growing."

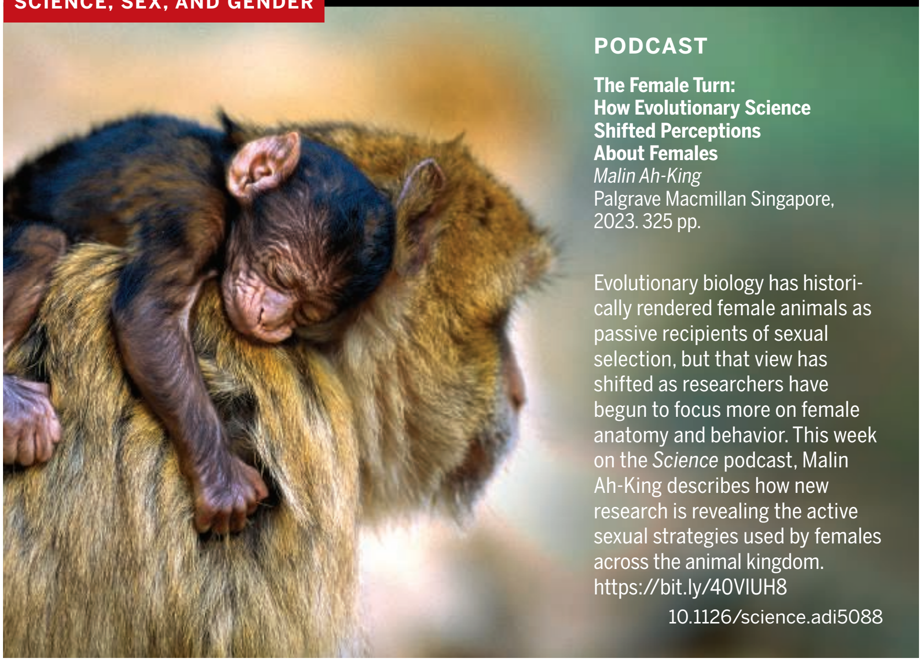
He also sees hope in Fort McMurray's recovery from the fire, even though ~20,000 residents never returned. He describes the revirescence of the burned city: "the forest around Fort Mac is slowly reviving…and the leveled neighborhoods are filling in with new homes, trees, and flowers."

Fire Weather is a gripping book that brings readers to the front lines of a major forest fire, while also exploring the intertwined history of oil and gas development and the study of climate change. Its lessons should not be soon forgotten. ■

10.1126/science.adh9238

The reviewer is a freelance science writer and editor and cofounder of *Science Borealis*. Email: snowhydro1@gmail.com

SCIENCE, SEX, AND GENDER



PODCAST

The Female Turn: How Evolutionary Science Shifted Perceptions About Females
Malin Ah-King
Palgrave Macmillan Singapore, 2023. 325 pp.

Evolutionary biology has historically rendered female animals as passive recipients of sexual selection, but that view has shifted as researchers have begun to focus more on female anatomy and behavior. This week on the *Science* podcast, Malin Ah-King describes how new research is revealing the active sexual strategies used by females across the animal kingdom. https://bit.ly/40VIUH8

10.1126/science.adi5088



LETTERS

The annualized real rate of return on 10-year treasury bonds is used to estimate the risk-free discount rate that US federal agencies should apply in benefit-cost analyses.

PHOTO: LARRY1235/SHUTTERSTOCK

Edited by **Jennifer Sills**

US benefit-cost analysis requires revision

Benefit-cost analysis should be guided by the best available data and methods. Yet when assessing regulations, US agencies currently value future impacts using outdated methods and data to determine discount rates. A proposal to revise a document known as Circular A-4 would address this deficiency by modernizing discounting guidance (*1*), which would substantially improve US regulatory analysis. A corresponding update has also been drafted for Circular A-94, which will improve US federal public investment analysis (*2*).

A social discount rate is used in benefit-cost analysis to translate impacts that occur at different times into comparable present values. Circular A-4, a US regulatory guidance document written in 2003 (*3*), directs agencies to use discount rates of 3% and 7% in most contexts. The 3% rate reflects the then-estimated risk-free rate at which society discounts consumption in the future. The 7% rate reflects the then-anticipated market rate of return to capital. These rates matter. For example, a hypothetical regulation might cost $20 billion today and yield $100 billion in undiscounted benefits in 30 years. Using the 7% discount rate, that $100 billion would be equivalent to $13 billion today. Regulators would compare the $20 billion investment with the $13 billion benefit and find a net loss of $7 billion. If a 3% rate is used instead, the same $20 billion investment, yielding the same $100 billion in 30 years, would be predicted to provide benefits equivalent to $41 billion today, a net gain of $21 billion. A lower discount rate thus provides agencies with more incentive to impose short-term costs to obtain long-term benefits such as environmental and health improvements.

Much has changed since 2003. Real returns to US Treasury notes—a common measure of the risk-free consumption rate—have trended lower, reflecting deep structural changes in the economy (*4*, *5*). Recent economic literature strongly supports the use of a consumption discount rate over a capital rate of return over longer time horizons (*6*, *7*) and shows that lower discount rates are appropriate for valuing long-term effects (*8*, *9*).

Consistent with these principles, the US Office of Management and Budget's (OMB's) proposed update (*1*) uses more recent economic data to lower the consumption-based discount rate from 3% to close to 2%. It also ends the use of the 7% discount rate in regulatory analysis, replacing capital-based discount rates with a shadow price of capital approach that bounds the potential impacts of public interventions on capital (*10*). In addition, the update calls for a Ramsey framework that links discounting to economic growth and a declining rate over longer time horizons (*11*).

OMB could improve its update further by improving its discussion of risk. The draft focuses on risk-free discount rates and recommends modeling the insurance value of a policy with uncertain net benefits using certainty-equivalents (*9*). However, agencies may lack sufficient expertise to make such calculations. OMB should offer additional guidance on how to convert cash flows to certainty-equivalents, including how to specify preference parameters consistent with the risk-free rate (*11*). To ease agency burden, OMB should provide an approximation of the average social risk premium adjustment (similar to the approach used in the draft Circular A-94) for discounting expected net benefits. Similarly, OMB should provide guidance on relative prices in valuing environmental services (*12*).

The proposed Circular A-4 update is open for public comment until 6 June. Experts in the field should support the update and suggest further revisions. Economists should also continue to improve discounting theory, especially the assessment of capital displacement and augmentation, risk premiums, and realistic preferences.

Peter H. Howard[1*], Max Sarinsky[1], Michael Bauer[2], Caroline Cecot[3], Maureen Cropper[4], Moritz Drupp[2], Mark Freeman[5], Kenneth T. Gillingham[6,7], Christian Gollier[8], Ben Groom[9], Qingran Li[10], Michael Livermore[11], Richard Newell[12], William A. Pizer[12,13], Brian Prest[12], Glenn Rudebusch[14,15], Thomas Sterner[16], Gernot Wagner[17]

[1]Institute for Policy Integrity, New York University School of Law, New York, NY, USA. [2]Universität Hamburg, Hamburg, Germany. [3]George Mason University, Fairfax, VA, USA. [4]University of Maryland, College Park, MD, USA. [5]University of York, York, UK. [6]Yale University, New Haven, CT, USA. [7]National Bureau of Economic Research, Cambridge, MA, USA. [8]Toulouse School of Economics, University of Toulouse-Capitole, Toulouse, France. [9]University of Exeter, Exeter, UK. [10]Clarkson University, Potsdam, NY, USA. [11]University of Virginia, Charlottesville, VA, USA. [12]Resources for the Future, Washington, DC, USA. [13]Duke University, Durham, NC, USA. [14]The Brookings Institution, Washington, DC, USA. [15]New York University, New York, NY, USA. [16]The University of Gothenburg, Gothenburg, Sweden. [17]Columbia University, New York, NY, USA.
*Corresponding author.
Email: peter.howard@nyu.edu

REFERENCES AND NOTES

1. Office of Management and Budget, "Circular A–4: Draft for Public Review" (2023).
2. Office of Management and Budget, "Circular A–94: Draft for Public Review" (2023).
3. Office of Management and Budget, "Circular A–4: Regulatory Analysis" (2003).
4. M. D. Bauer, G. D. Rudebusch, *Rev. Econ. Stat*. 10.1162/rest_a_01109 (2021).
5. P. Howard, J. Schwartz, *Yale J. Regulation* **39**, 595 (2022).
6. M. A. Drupp, M. C. Freeman, B. Groom, F. Nesje, *Am. Econ. J. Econ. Pol.* **10**, 109 (2018).
7. Q. Li, W. A. Pizer, *J. Environ. Econ. Manag.* **107**, 102428 (2021).
8. K. Arrow *et al*., *Science* **341**, 349 (2013).
9. C. Gollier, *Am. Econ. Rev.* **104**, 534 (2014).
10. G. Newell, W. A. Pizer, B. C. Prest, "The shadow price of capital: Accounting for capital displacement in benefit-cost analysis" (Issue Brief 22-08, Resources for the Future, 2023).
11. R. G. Newell, W. A. Pizer, B. C. Prest, *J. Assoc. Environ. Resour. Econ.* **9**, 1017 (2022).
12. T. Sterner, U. M. Persson, *Rev. Environ. Econ. Pol.* **2**, 61 (2008).

COMPETING INTERESTS

C.C. is an elected, unpaid board member of the Society for Benefit-Cost Analysis.

10.1126/science.adi5943

US "China initiatives" promote racial bias

On behalf of the Society of Chinese Bioscientists in America, the Asian American Scholar Forum, the Chinese American Hematologist and Oncologist Network, and the Chinese Biological Investigators Society, we applaud the News Feature "Pall of suspicion" (J. Mervis, 24 March, p. 1180) and accompanying Editorial, "Eroding trust and collaboration" (H. H. Thorp, 24 March, p. 1171). These articles revealed secretive and widespread investigations of Chinese American scholars prompted by the National Institutes of Health (NIH) under what Mervis described as "NIH's China Initiative." As Mervis explains, "81% of the scientists cited in the NIH letters identify as Asian," and hundreds of Chinese American scholars' lives and careers have been disrupted or ruined. The Department of Justice's original "China Initiative," launched in 2018 to protect national security, raised concerns that Asian Americans and immigrants were being profiled. Although there have since been positive changes to that program (*1*), both scientists and nonscientists in the United States must remain vigilant to ensure that history does not repeat itself.

Tensions between the United States and China are likely to increase, but Chinese Americans should not be treated as collateral damage. Systems that promote bigotry against individuals of any ethnic background should not be tolerated and have no place within the US government. The United States, as a leader in science and technology, must adhere to the principles that foster a culture of inclusion, diversity, and equity. This focus will help attract the best and brightest talents from abroad, including China. The NIH policies described in the News story have negatively affected Asian Americans and eroded US leadership in science and technology.

We urge the federal government to shift away from deterrence and fear and instead focus on the philosophies that have made the United States a global leader in science and technology: openness, diversity, and international collaboration. We also call on federal agencies to work with the Asian American scholar community and academic institutions toward solutions that allow the United States to retain and attract talent as well as provide a safe and welcoming academic environment. Transparency, due process, and accountability can address biases and rebuild trust with Asian American and scholar communities.

Members of the Asian American community have experienced years of not only government suspicion but also anti-Asian hate and violence (*2*). The US government and academic institutions should provide support during these times of increased anti-Asian sentiments. The members of our organizations have faith that the country will live up to American ideals.

Shan-Lu Liu[1]*, Lishan Su[2], Kunxin Luo[3], Kai Li[4], Gang Chen[5], Xiaodong Zhang[5], Bo Zhao[6], RuiRong Yuan[7], Yingzi Yang[8], Lee Zou[9], Chuan He[10,11], Jing Yang[12], Lin He[13], Yi Li[13], Dong Wang[13], Zhigang Suo[5], Gisela Perez Kusakawa[14], Yasheng Huang[15]

[1]President, Society of Chinese Bioscientists in America (SCBA), Boston, MA, USA. [2]President-Elect, SCBA, Boston, MA, USA. [3]Secretary, SCBA, Boston, MA, USA. [4]Vice President, Asian American Scholar Forum (AASF), New York, NY, USA. [5]Board Member, AASF, New York, NY, USA. [6]President, Chinese American Hematologist and Oncologist Network (CAHON), Milwaukee, WI, USA. [7]Board Chair, CAHON, Milwaukee, WI, USA. [8]President, Chinese Biological Investigators Society (CBIS), Brookline, MA, USA. [9]President-Elect, CBIS, Brookline, MA, USA. [10]Executive Secretary, CBIS, Brookline, MA, USA. [11]Treasurer, AASF, New York, NY, USA. [12]Treasurer, CBIS, Brookline, MA, USA. [13]Board Member, CBIS, Brookline, MA, USA. [14]Executive Director, AASF, New York, NY, USA. [15]President, AASF, New York, NY, USA. *Corresponding author. Email: liushan1115@gmail.com

The views expressed in this Letter are solely those of the authors and the four organizations, not the affiliated institutions of the authors.

REFERENCES AND NOTES

1. US Department of Justice, "Assistant Attorney General Matthew Olsen delivers remarks on countering nation-state threats" (2022); https://www.justice.gov/opa/speech/assistant-attorney-general-matthew-olsen-delivers-remarks-countering-nation-state-threats.
2. T. Tang, "More than 9,000 Anti-Asian incidents have been reported since the pandemic began" *The Associated Press* (2021).

COMPETING INTERESTS

The US Department of Justice charged G.C. as part of its China Initiative in January 2021 and dropped all charges in January 2022.

10.1126/science.adi4909

Safeguarding integrity and collaborations

In his News Feature "Pall of suspicion" (24 March, p. 1180), J. Mervis describes 5 of the 246 cases in which the US National Institutes of Health (NIH) suspected that scientists were violating grant funding policies. Mervis insinuates that the cases were mishandled and unjustified, but the data he cites show that NIH had well-founded concerns. In over 60% of the 246 cases, institutional investigations resulted in employment termination or future grant ineligibility, whereas institutions found no evidence of wrongdoing in only 6% of cases. Mervis characterizes the investigations as secretive, but the lack of transparency can be explained by privacy considerations, which prevented NIH from disclosing details about the evidence leading to concerns about these 246 scientists. Although nondisclosure of details to the public may have led to misunderstandings among those not directly involved, NIH followed appropriate procedures and guidelines to protect the integrity of its funding.

In 2016, the Federal Bureau of Investigation informed NIH of possible breaches of peer review integrity—grant reviewers were violating confidentiality attestations by sharing applications outside of peer review meetings. Over the next 2 years, NIH learned about dozens of scientists in the United States potentially involved with Chinese talent recruitment programs, a mechanism that the Chinese government uses to recruit and employ experts in science and technology, often without informing their American universities. In 2018, NIH Director Francis Collins informed institutions that the NIH Office of Extramural Research might contact them regarding specific cases, most of which were identified through institutional self-disclosures or discrepancies between grant records and publications. NIH began this work well before the 2018 launch of the "China Initiative," a separate effort led by the US Department of Justice aimed at curtailing Chinese espionage in science and technology.

NIH has publicized its policies and procedures, as well as aggregate data on its compliance reviews of scientists (*1*). Allegations involved more than 600 scientists, of which NIH deemed 246 serious enough to pursue (*1*). Cases have revealed concerning patterns of deceptive behavior, which are not victimless; given NIH's competitive grant system, dishonest scientists deny funding to those who act with integrity. However, the 232 scientists who were not cleared by compliance reviews make up less than 1% of all NIH-funded principal investigators (*2*). And since 2020, the number of new cases of alleged grant policy violations has decreased substantially (*1*), perhaps due to concerted efforts by professional groups and institutions to enhance oversight and internal controls.

Cross-government efforts have supported grant oversight. NIH worked with the Office of Science and Technology Policy on National Security Presidential Memorandum 33 implementation (*3*), partnered with other agencies on the development of research security programs, participated in National Academies and international programs, and has been responsive to audits and Congressional oversight. In Fiscal Years 2022 and 2023, Congress allocated $2.5 million to NIH work on foreign interference cases (*4*).

Many NIH cases of grant policy

violations involve activities in China and scientists of Chinese origin, raising concerns about racial, ethnic, or political targeting. The disproportionate share may be related to the extensive reach of Chinese talent recruitment programs, which explicitly state preferences for ethnic Chinese scientists (*5*). NIH compliance reviews abide by the same procedures for all scientists irrespective of their demographics, and the inquiries include cases involving countries other than China (*1*).

NIH does not have to choose between rigorous integrity oversight and international collaboration. Proper collaborations do not entail stealth employment, duplicative funding, undisclosed financial conflicts of interest, or the repeated recitation of lies to institutional or government officials. Moreover, dishonest behaviors lead stakeholders outside academia (*6*) to question whether NIH is a responsible steward of taxpayer dollars.

Constructive steps have been taken to facilitate international collaboration. In 2019, the JASON report, commissioned by the National Science Foundation, proposed a set of questions that scientists and institutions should consider when establishing relationships with colleagues abroad (*7*). Scientists should make sure that terms of engagement are made clear in writing and that all participants (including scientists' domestic employers) are fully informed. Participants should address conflicts of interest, conflicts of commitment, funding, resource and data sharing, reporting requirements, travel expectations, dissemination of results, and engagement termination. Institutional leaders should consider political, civil, and human rights risks; intellectual property; export controls; data and publication; misrepresentation risks; risks to community and core values; and risks of not engaging. Recently, leaders at the Massachusetts Institute of Technology identified lines that should not be crossed, such as participation in Chinese talent recruitment programs that are designed to transfer technology to China (*6*). Together, stakeholders in diverse sectors should strive to envision a framework that secures trust and integrity while enhancing all that robust international collaborations have to offer.

Michael S. Lauer* and Patricia A. Valdez
Office of Extramural Research, Office of the Director, National Institutes of Health, Bethesda, MD 20892, USA.
*Corresponding author. Email: lauerm@nih.gov

REFERENCES AND NOTES

1. National Institutes of Health Foreign Interference (2022); https://grants.nih.gov/policy/foreign-interference/data.
2. M. Lauer, "Inequalities in the distribution of National Institutes of Health research project grant funding" (2022), Table 5; https://nexus.od.nih.gov/all/2022/01/18/inequalities-in-the-distribution-of-national-institutes-of-health-research-project-grant-funding/. The first row in Table 5 shows that 34,936 PIs received grants in FY2020. NIH identified serious violations perpetrated by 208 investigators (*1*), representing about 0.6% of the total recipients.
3. National Science and Technology Council, "Guidance for implementing National Security Presidential Memorandum 33 (NSPM-33) on national security for the United States Government-supported research and development, a report by the Subcommittee on Research Security, Joint Committee on the Research Environment" (2022); https://www.whitehouse.gov/wp-content/uploads/2022/01/010422-NSPM-33-Implementation-Guidance.pdf.
4. Division H—Departments Of Labor, Health, and Human Services, and Education, and Related Agencies, "Appropriations Act, 2023" (2023), p. 88; https://www.appropriations.senate.gov/imo/media/doc/Division%20H%20-%20LHHS%20Statement%20FY23.pdf.
5. 2019 Chinese Academy of Sciences Talent Program Application Guide (2019); https://cset.georgetown.edu/wp-content/uploads/t0151_CAS_talent_program_guide_EN.pdf.
6. R. Lester *et al.*, *Science* **380**, 246 (2023).
7. JASON, "Fundamental research security" (2019); https://www.nsf.gov/news/special_reports/jasonsecurity/JSR-19-2IFundamentalResearchSecurity_12062019FINAL.pdf.

10.1126/science.adi3894



RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



BIOGEOGRAPHY

It's the rocks

Biodiversity is typically high in mountain ranges, likely because of their history of tectonic activity and wide range of environments from base to peak. Looking closer, however, reveals that mountain diversity often occurs within environments and in ranges that have long ceased being tectonically active. Stokes *et al*. describe diversification among lineages of a freshwater fish species in the North American Appalachian Mountains driven by an affiliation with metamorphic rock. Specifically, as metamorphic rock wore away and was replaced by other types, Greenfin Darters were increasingly isolated in upstream reaches where metamorphic rock remained, leading to population separation and allopatric diversification. —SNV

Science, add9791, this issue p. 855

Populations of Greenfin Darters, pictured above, have become isolated in mountain areas as their habitat range has fragmented due to erosion.

HUMAN GENETICS

Quebec landscape shapes populations

Some of the earliest population genetic studies in humans revealed clines of genetic variation across Europe and other regions, showing isolation by distance on broad scales. Following this tradition, Anderson-Trocmé *et al*. used spatial data coupled with pedigrees and genetic samples to show how geographic barriers have affected French Canadian migration and settlement in Quebec since the 17th century. The authors simulated 1.4 million genomes representing modern-day individuals living in this region, in addition to their ancestors. These analyses and this resource together reveal insights into the historic migration and settlement patterns of this population, which were shaped by geographic constraints. —CNS

Science, add5300, this issue p. 849

OCEAN MICROBIOLOGY

Insight into cyanobacteria aggregates

The nitrogen-fixing cyanobacterium *Trichodesmium* forms filamentous aggregates in response to stress. Pfreundt *et al*. subjected *Trichodesmium* cultures to excess light and reactive oxygen species, which encouraged them to form aggregates. Analysis of images and videos of individuals and pairs of filaments revealed how their motion affects overall colony architecture. If long segments of filament pairs lose contact with one another, then a change in the direction of their motion ensures that the filaments remain attached. Mathematical modeling demonstrated that responsive filament motion acts as a positive feedback mechanism to maintain aggregate architecture. —MRS

Science, adf2753, this issue p. 830

CELL BIOLOGY

Receptor shedding to stop signaling

Cytotoxic T lymphocytes (CTLs) are serial killers that can destroy multiple target cells in succession. Using high-resolution three-dimensional imaging, Stinchcombe *et al*. found that T cell receptor (TCR) activation causes rapid membrane specialization at the immune synapse so that membrane containing activated TCRs buds off in a process known as "ectocytosis." As ectosomes bud off, CTLs and their target cells separate, allowing serial killing to proceed. Thus, in activated CTLs, ectocytosis renders TCR signaling self-limiting. —SMH

Science, abp8933, this issue p. 818

OCEAN NUTRIENTS

Venting iron

Large areas of the world's surface ocean contain limited

concentrations of iron, most of which is supplied by upwelling of deeper, nutrient-rich water or atmospheric deposition. Bonnet *et al*. found that shallow hydrothermal sources also can supply iron to the surface ocean. They show that fluids emitted along the Tonga volcanic arc in the South Pacific supply enough iron to the photic layer to stimulate biological activity well above that in adjacent unfertilized waters. —HJS

Science, abq4654, this issue p. 812

2D MATERIALS

Correlated excitons

Moiré multilayers of graphene and transition metal dichalcogenides have been found to host a number of exotic electronic phases. In addition to charge carriers—electrons and holes—these two-dimensional structures support their pairs, called excitons, which are also expected to form correlated phases. Xiong *et al*. observed a correlated insulator of interlayer excitons in a bilayer consisting of tungsten diselenide, which hosted holes, and tungsten disulfide, which hosted electrons. The researchers measured the optical response of the material to excitation by light and found signatures of an incompressible state of excitons. —JS

Science, add5574, this issue p. 860

HIV

Activating T cells against HIV-1

An effective vaccine for HIV-1 remains an essential but elusive goal. One promising HIV-1 vaccine candidate, eOD-GT8, has previously been shown to elicit HIV-specific B cell and antibody responses in humans. Cohen *et al*. investigated the extent to which T cell responses were elicited in these vaccine recipients. Polyfunctional helper T cell responses were elicited after vaccination and targeted both the eOD-GT8 and the lumazine synthase components of the vaccine. These T cell responses may enhance vaccine efficacy, including after subsequent booster vaccines. —CSM

Sci. Transl. Med. (2023) 10.1126/scitranslmed.adf3309

PHYSIOLOGY

Maintaining a quiescent endothelium

Atherosclerosis causes disturbed blood flow that activates inflammatory pathways in endothelial cells, which recruits leukocytes to plaques and exacerbates disease progression. Bosseboeuf *et al.* investigated the role of the transmembrane protein NRP1 in the response of the endothelium to flow. Under normal flow patterns, NRP1 stabilized protein complexes at cell-cell junctions and suppressed endothelial inflammation. In mice, NRP1 deficiency was associated with greater numbers of rolling leukocytes on endothelial cells and larger plaque sizes in a model of atherosclerosis. —WW

Sci. Signal. (2023) 10.1126/scisignal.abo4863

SOLAR CELLS

Chelators for perovskite solar cells

Although small-area n-i-p structure perovskite solar cells can have high efficiency, commercialization will require large-area p-i-n structures with highly uniform perovskites and charge-transport layers. Fei *et al*. found that bathocuproine, which is often used as an electron-transport material, can improve power-conversion efficiency and stability when added to the hole-transport layer. The chelation product of bathocuproine with lead ions is insoluble in the perovskite ink and also decreases the formation of amorphous regions by reducing the amount of trapped dimethyl sulfoxide solvent. Minimodules with an aperture area of 26.9 square centimeters had a certified efficiency of 21.8% and light-soaking stability exceeding 2000 hours. —PDS

Science, ade9463, this issue p. 823

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



PHOTO: EYEPRESS/NEWSCOM

MARINE POLLUTION

Accelerating accident appraisal

In May 2021, a new container ship, the *X-Press Pearl*, caught fire and burned for 12 days. It finally sank in the coastal waters of Sri Lanka, releasing vast quantities of plastic pellets and a slurry of toxic compounds. Zhang *et al.* highlight multiple risky practices in the container ship industry, including failure to declare toxic items within containers, as well as poor packing and stacking, the ever-increasing size of ships, and container loss from increased storm activity. In a companion paper, the same authors argue for more timely biological damage assessment to determine the potential for harm caused by such accidents. They propose using whole-cell bioreporters that not only identify the presence of toxins but also gauge the biotic stress that they may cause within the affected environment. —SNV

Mar. Poll. Bull. (2023) 10.1016/j.marpolbul.2023.114728, 10.1016/j.marpolbul.2023.114705

SYMBIOSIS

Cells containing multitudes

Endosymbiosis occurs when an organism becomes resident within the cells of another organism and contributes to the metabolism of the host. For example, chloroplasts, which conduct photosynthesis in plant and algal cells, may have evolved from endosymbiotic cyanobacteria. Some cells host multiple endosymbionts. George *et al.* found that some cryptomonad algae (e.g., *Cryptomonas gyropyrenoidosa*) have as many as seven types of endosymbionts, including mitochondria, plastids, and nucleomorphs, as well as two intracellular bacteria, *Megaira polyxenophila* and *Grellia*

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

MEDICINE

Making regenerative medicine a reality

Restoring health after tissue damage caused by infection, injury, or disease requires tissue regeneration. This area of medical research involves approaches to promote tissue regeneration and to suppress some natural processes that can limit tissue function. In a Perspective, McKinley *et al.* address the current approaches and challenges of regenerative medicine. They discuss stem cell grafts and tissue reprogramming to promote tissue restoration, as well as recent efforts to modulate inflammation and fibrosis. As these strategies advance through preclinical research and start to enter clinical trials, the possibility of regenerative therapies seems closer, despite numerous challenges ahead. —GKA

Science, add6492, this issue p. 796

NEUROSCIENCE

Synaptic proteasome-regulatory particles

Some important cellular functions, such as protein synthesis and degradation, are performed by nanoscale machines built of multisubunit protein complexes. These biological machines are assembled from individual components, and any excess or unassembled proteins are typically rapidly degraded. The proteasome is the major protein degradation machine in cells and is assembled from subcomplexes, one regulatory and one catalytic. Sun *et al*. studied the distribution of proteasomes and their subcomplexes in neurons (see the Perspective by Türker and Margolis). In neuronal dendrites, the regulatory subcomplex of the proteasome was found in excess and in isolation from its partner catalytic subcomplex. The isolated subcomplex helped to regulate synaptic proteins and synaptic transmission. —SMH

Science, adf2018, this issue p. 811;
see also adi1166, p. 795

THERMOELECTRICS

Filling vacancies

Thermoelectric materials interconvert heat and electricity, making them useful for a range of devices. Liu *et al*. added copper to tin selenide, which improved the thermoelectric and mechanical properties near room temperature (see the Perspective by Chung). Tin selenide tends to have several defects when synthesized, including tin vacancies. The copper occupies these intrinsic tin vacancies, leading to improved carrier mobility. The overall strategy creates a lattice less riddled with vacancies, which could be useful for other materials. —BG

Science, adg7196, this issue p. 841;
see also adi2174, p. 800

HIBERNATION

Breeding mismatch?

The Arctic is warming faster than other parts of the world, with potential major impacts on the cold-adapted species that live there. Using a 25-year dataset on arctic ground squirrels, Chmura *et al*. found that the slower freezing of the permafrost has led to these squirrels delaying the timing of heat production during hibernation. More worrying, however, is that over the past two decades, females have begun to emerge from hibernation earlier, whereas males have not. Such a mismatch in emergence could affect reproduction. These types of accumulating responses to climate change, although seemingly small, can have cascading effects on species and systems. —SNV

Science, adf5341, this issue p. 846

PLANT ECOLOGY

Microbes help trees adjust to climate

As sessile organisms, plants have limited ability to move in response to climate change. However, the microbial communities that help plants access soil nutrients may also increase their tolerance to climate stress. Using field and laboratory experiments, Allsup *et al*. showed that soil microbe communities sourced from colder or drier sites increased the survival of inoculated tree seedlings exposed to cold winters or drought, respectively (see the Perspective by Afkhami). These findings suggest greater potential resilience for trees and highlight the importance of species interactions in determining their response to climate change. —BEL

Science, adf2027, this issue p. 835;
see also adi1594, p. 798

HIV

Spontaneous controller secrets revealed

Individuals who spontaneously control HIV infection without antiviral treatment have provided critical insights into natural HIV immunity. Collins *et al*. studied $CD8^+$ T cells in spontaneous controllers of HIV to better understand the mechanisms involved in HIV control. The authors examined the phenotypic, transcriptomic, and functional features; spatial localization; and clonotypic compartmentalization of peripheral and lymph node (LN)–resident CD8 T cells. Controllers displayed antigen-specific proliferation and cytolytic potential that were different from noncontrollers. Within controllers, T cell receptor analysis confirmed shared clonotypes detected in both peripheral and LN $CD8^+$ T cells, although the latter had a gene signature enriched in effector and inflammatory chemotaxis functions. $CXCR5^+$ follicular $CD8^+$ T cells enriched in controllers expressed cytolytic effectors near HIV RNA foci in germinal centers. These findings provide insight into how HIV-specific lymphoid cells mediate control of HIV in LNs. —CNF

Sci. Immunol. (2023)
10.1126/sciimmunol.ade5872

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

An abundance of free regulatory (19*S*) proteasome particles regulates neuronal synapses

Chao Sun, Kristina Desch†, Belquis Nassim-Assir†, Stefano L. Giandomenico†, Paulina Nemcova, Julian D. Langer, Erin M. Schuman*

INTRODUCTION: An individual neuron hosts about 10^4 synapses in its expansive dendritic and axonal arbor. To support synaptic transmission and plasticity, both protein synthesis and degradation machines are localized to remodel individual synaptic proteomes. The machines that are responsible for most cytosolic protein degradation (proteasomes) are assembled from the catalytic core (20*S*) and the regulatory particle (19*S*), which associates with the 20*S* at one (26*S*) or both ends (30*S*). The 19*S* regulatory particle recognizes and deubiquitylates substrates as they enter the 20*S* catalytic core. Here, we explored the subcellular abundance of the 20*S* and 19*S* particles and discovered an independent function for the 19*S* proteasome in the regulation of synapses.

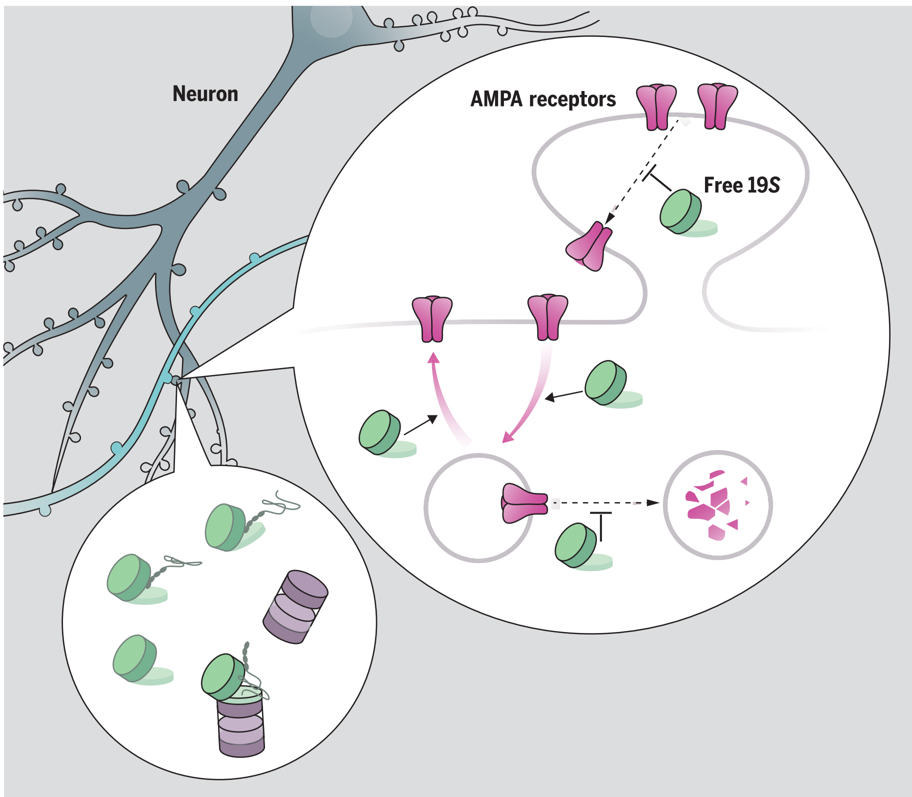
RATIONALE: The average half-life of the proteins that make up the neuronal proteasome-regulatory subcomplex (19*S*) is significantly different from those that comprise the catalytic subcomplex (20*S*), suggesting that the two subcomplexes do not always function in concert. Indeed, emerging evidence points to distinct functions of free 20*S* particles. These data suggest that the 26*S* proteasome is not the only assembly or functional state of synaptic proteasome particles.

RESULTS: To visualize proteasome 20*S* and 19*S* subcomplexes in neuronal subcellular compartments, we conducted immunolabeling for 19*S* and 20*S* protein subunits in mouse brain slices and in cultured rat cortical neurons using confocal microscopy. We found that the 19*S* and 20*S* proteasome particles were differentially distributed in neuronal subcellular compartments, raising the possibility of different proteasome formats at synapses.

To elucidate further the abundance and assembly states of 19*S* and 20*S*, we visualized dendritic 19*S* and 20*S* particles using quantitative, multiplexed, single-molecule localization microscopy (DNA-PAINT) and found an abundance of free 19*S* particles in dendrites and synapses. Furthermore, we found that lysine 63 (Lys^{63})–ubiquitylated proteins were predominantly associated with the free 19*S* particle rather than with the 26*S* or 30*S* particles in neuronal lysates. Indeed, free 19*S* particles could act as a proteasome-independent deubiquitylase (DUB) against Lys^{63} ubiquitylation in vitro.

To investigate the potential neuronal substrates, we performed immunoprecipitation of Lys^{63}-ubiquitylated proteins and found that synaptic proteins were major interactors of Lys^{63}-linked ubiquitin chains (Lys^{63}-ub). Pharmacological inhibition of the free 19*S* DUB activity altered both miniature and evoked synaptic transmission. Consistent with this, other experiments showed that the free 19*S* serves as a multipronged regulator that maximizes the synaptic availability of AMPA receptors (AMPARs).

CONCLUSION: Up to 90% of the total proteome of a single neuron resides in dendrites and axons. Local proteasomal function is thus essential for synaptic homeostasis as well as plasticity. Here, we found an unexpected abundance of free 19*S* regulatory particles near synapses. We discovered that these free 19*S* particles deubiquitylate important Lys^{63}-ubiquitylated synaptic proteins such as AMPARs. The dynamics of Lys^{63} ubiquitylation and deubiquitylation potently regulated synaptic transmission. The excess of the free 19*S* proteasome particles present in dendrites thus represents an interesting example of a subcomplex of a multiprotein machine performing a moonlighting function near synapses, in this case as an independent "editor" of a synaptic ubiquitin code. ■



Moonlighting function of the proteasome-regulatory subcomplex near synapses. Shown is a neuron (gray) with its dendritic spines that contact axons (teal) to form synapses. At synapses, quantitative measurements of the different proteasome components revealed an abundance of free 19*S* (green) subunits that can recognize ubiquitylated protein substrates. The abundant free 19*S* subunits can bind and deubiquitylate the non-proteasome-targeting Lys^{63}-ub associated with synaptic proteins. Manipulation of 19*S* deubiquitylase activity substantially altered excitatory synaptic transmission and regulated the synaptic availability of AMPARs at multiple trafficking points (synaptic dwell time, endo- and exocytosis, and lysosome targeting) in a proteasome-independent manner.

The list of author affiliations is available in the full article online.
***Corresponding author. Email: erin.schuman@brain.mpg.de**
†These authors contributed equally to this work.
Cite this article as C. Sun *et al.*, *Science* 380, eadf2018 (2023). DOI: 10.1126/science.adf2018

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adf2018

RESEARCH ARTICLE

NEUROSCIENCE

An abundance of free regulatory (19*S*) proteasome particles regulates neuronal synapses

Chao Sun[1]†‡, Kristina Desch[1]§, Belquis Nassim-Assir[1]§, Stefano L. Giandomenico[1]§, Paulina Nemcova[1]¶, Julian D. Langer[1,2], Erin M. Schuman[1]*

The proteasome, the major protein-degradation machine in cells, regulates neuronal synapses and long-term information storage. Here, using super-resolution microscopy, we found that the two essential subcomplexes of the proteasome, the regulatory (19*S*) and catalytic (20*S*) particles, are differentially distributed within individual rat cortical neurons. We discovered an unexpected abundance of free 19*S* particles near synapses. The free neuronal 19*S* particles bind and deubiquitylate lysine 63–ubiquitin (Lys[63]-ub), a non–proteasome-targeting ubiquitin linkage. Pull-down assays revealed a significant overrepresentation of synaptic molecules as Lys[63]-ub interactors. Inhibition of the 19*S* deubiquitylase activity significantly altered excitatory synaptic transmission and reduced the synaptic availability of AMPA receptors at multiple trafficking points in a proteasome-independent manner. Together, these results reveal a moonlighting function of the regulatory proteasomal subcomplex near synapses.

An individual neuron hosts ~10^4 synapses in its expansive dendritic and axonal arbor (*1*). To support parallel synaptic computation and plasticity, both protein synthesis and degradation machines are localized to remodel individual synaptic proteomes on site (*2*–*6*). The machines that perform these functions, the ribosomes and proteasomes, are assembled from multiple functional subunits, giving rise to different assembly formats near synapses (*7*–*10*). Recent work has suggested that these different assembly formats can exhibit distinct functional preferences within neuronal subcellular compartments. For example, although polyribosomes (three or more ribosomes engaged with an mRNA) have been regarded as the main form of protein-synthesis machinery (*11*), single ribosomes (monosomes) targeting a distinct set of synaptic mRNAs have recently been described as an important source of translation near synapses (*7*). Data such as these raise the possibility that neurons can use alternative formats of machines to perform different functions akin to protein "moonlighting." The proteasome is a multiprotein complex that is responsible for most cytosolic protein degradation in mammalian cells (*12*); it operates at synapses (*6*, *13*), and its function is necessary for long-term memory (*14*, *15*). Proteasomes consist of regulatory (19*S*) and catalytic (20*S*) particles (*16*). Proteins are targeted for proteasomal degradation by the attachment of ubiquitin chains that are primarily linked at lysine 48 (Lys[48]) (*17*). However, there is another major type of ubiquitylation in mammalian cells, ubiquitin chains linked through lysine 63 (Lys[63]) ubiquitylation (*18*). Lys[63]-ubiquitylated proteins are not associated with proteasomes (*19*). The process of proteasomal degradation includes the Lys[48] deubiquitylation of protein targets by the 19*S* particle, followed by the protease-mediated degradation within the barrel-shaped 20*S* particle (*16*). Thus, canonical proteasome function requires the stable assembly of the regulatory 19*S* and the catalytic 20*S* particles (*20*, *21*). As might be expected, the average expression levels of 19*S* and 20*S* protein subunits are also stoichiometric for singly capped proteasomes (26*S*) consisting of one 19*S* particle and one 20*S* particle (*22*). Consistent with this subunit expression, a cryogenic electron tomography study detected capped neuronal proteasomes primarily as singly capped 26*S* proteasomes with a minor fraction of 30*S* doubly capped proteasomes (*23*). In addition, neuronal synapses sequester both types of proteasomal particles (19*S* and 20*S*) in an activity-dependent manner (*6*).

The average half-life of the proteins that make up the neuronal 19*S* particle, however, is significantly different from those that comprise the 20*S* particle, suggesting that the two subcomplexes do not always turn over in concert (*24*). In addition, emerging evidence also points to distinct functions of free 20*S* particles (*25*, *26*), the detection of which remains elusive in cryo–electron tomography. These data suggest that the 26*S* proteasome is not the only assembly or functional state of synaptic proteasome particles. Considering the huge protein turnover demand in neuronal processes (*3*), the nature of synaptic proteasomal assembly states represents a major knowledge gap. Here, we show first that neuronal proteasome regulatory and catalytic particles are unevenly distributed within neurons both in vivo and in vitro. An "excess" of proteasome-regulatory particles 19*S* (relative to 20*S*) was detected near synapses. Combining single-molecule localization microscopy and native protein gel electrophoresis, we profiled the dendritic proteasome assembly states and revealed a moonlighting function for free 19*S* proteasome regulatory particles as deubiquitylases (DUBs) against Lys[63]-ub chains that target synaptic proteins. Excess proteasome-regulatory particles thus modify an unconventional form of ubiquitylation and regulate crucial synaptic proteins in parallel with their conventional role in proteasome degradation.

Results

An abundance of the free 19S proteasome near synapses

To investigate the subcellular distribution of the 19*S* proteasome-regulatory particles and 20*S* catalytic particles, we conducted immunolabeling for 19*S* and 20*S* protein subunits in mouse brain slices and in cultured rat cortical neurons using confocal microscopy. Because the 20*S* particle contains two copies of each constituent protein subunit, whereas the 19*S* particle contains only one copy of each protein subunit (Fig. 1A) (*16*), we targeted the two copies of PSMA7 subunit in each 20*S* particle and two subunits of each 19*S* particle (PSMC1 and PSMD6) using validated antibodies (fig. S1, A to E). If the 26*S* proteasome is the main assembly state of both neuronal proteasomal particles (Fig. 1A), then the 19*S* and 20*S* immunofluorescence should distribute in a consistent ratio in neuronal cell bodies and their processes. However, this was not what we observed. The relative fluorescence of the 19*S* and 20*S* subunits varied in different subcellular compartments of cultured rat cortical neurons (Fig. 1B). The 19*S* subunit was significantly more abundant in dendrites and axons, whereas the 20*S* subunit was more abundant in the cell body (Fig. 1C). Antibodies targeting different proteasome subunits generated similar distribution patterns in cultured rat hippocampal neurons (fig. S1G). Consistent with this, in hippocampal area CA1 of coronal mouse brain slices, where neuronal cell bodies and processes exhibit a clear laminar organization (Fig. 1D and fig. S1F), the 20*S* fluorescence was relatively enriched in the somata layer (stratum pyramidale), whereas the 19*S* fluorescence was enriched in the neuropil layer (stratum radiatum). These observations suggest that the 19*S*

[1]Max Planck Institute for Brain Research, 60438 Frankfurt am Main, Germany. [2]Max Planck Institute for Biophysics, 60438 Frankfurt am Main, Germany.
*Corresponding author. Email: erin.schuman@brain.mpg.de
†Present address: Danish Research Institute of Translational Neuroscience (DANDRITE), Nordic-EMBL Partnership for Molecular Medicine, 8000 Aarhus C, Denmark.
‡Present address: Aarhus University, Department of Molecular Biology and Genetics, 8000 Aarhus C, Denmark.
§These authors contributed equally to this work.
¶Present address: Department of Neurosurgery, University Hospital Bonn, 53127 Bonn, Germany.

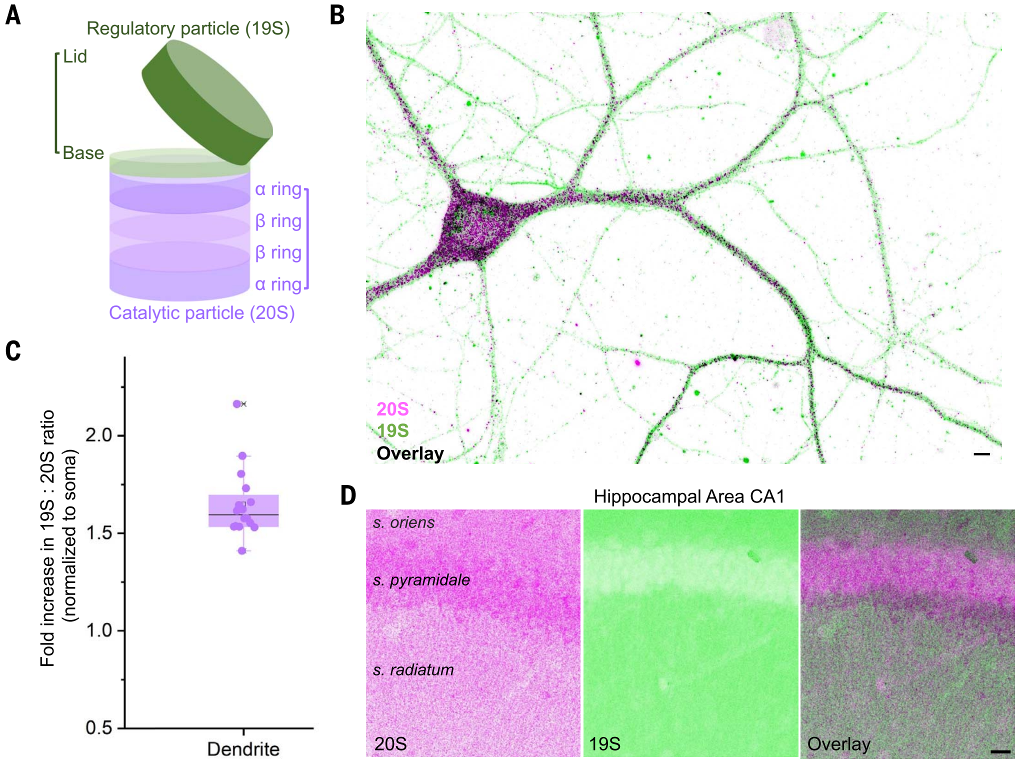


Fig. 1. Proteasome regulatory (19*S*) and catalytic (20*S*) particles are differentially distributed in neuronal subcellular compartments. (**A**) Illustration of a singly capped proteasome (26*S*) in which each 26*S* contains one copy of each 19*S* subunit (green) and two copies of each 20*S* subunit (magenta). (**B**) Airy-scanning confocal micrograph of a mature rat primary cortical neuron immunolabeled for the 19*S* (green; anti-PSMC1/PSMD6) and 20*S* (magenta; anti-PSMA7) particles. Scale bar, 4 μm. (**C**) Scatterplot showing an elevated ratio of 19*S*:20*S* fluorescence intensity (per unit area) in dendrites (>1.5; 17 dendritic branches) compared with somata (normalized to 1.0; six cell bodies). (**D**) Immunofluorescence staining of hippocampal area CA1 in a coronal section of the mouse brain (20*S* in magenta by anti-PSMA7, 19*S* in green by anti-PSMC1/PSMD6) showing the enriched 19*S* staining (green) in the stratum radiatum (neuropil) and enriched 20*S* staining (magenta) in the stratum pyramidale (somata). Scale bar, 25 μm.

and 20*S* proteasome particles are differentially distributed in neuronal subcellular compartments, raising the possibility of different proteasome formats at synapses.

To elucidate further the abundance and assembly states of 19*S* and 20*S* near synapses, we visualized dendritic 19*S* and 20*S* particles using quantitative, multiplexed, single-molecule localization microscopy (DNA-PAINT; see the materials and methods) (*5*, *27*). The 19*S* and 20*S* particles were detected as individual clusters of localizations at single-molecule resolution. We detected capped proteasomes as the colocalization between 19*S* and 20*S* particles (Fig. 2A; also see the materials and methods). On average, 2.5 copies of capped (19*S* and 20*S*) proteasome particles were found per micrometer of dendrite (Fig. 2B), with ~80% of the detected capped proteasomes containing one 19*S* and one 20*S* (e.g., the singly capped proteasome 26*S*) and ~20% containing one 20*S* and two 19*S* copies (e.g., the doubly capped proteasome 30*S*). The detected ratio between 26*S* and 30*S* proteasomes was similar to what was observed by a recent in situ cryogenic electron tomography study in neurons (*23*). Although the 26*S* was the main form of capped proteasome detected, free 19*S* and 20*S* particles were also abundant in neuronal dendrites (Fig. 2C). The ratio of total (free and capped) 19*S* to 20*S* particles in dendrites was ~2:1 (Fig. 2D), exceeding the expected 1:1 ratio needed to assemble the 26*S* proteasome. Despite the superstoichiometric presence of the 19*S* in dendrites, nearly half of the 20*S* particles remained uncapped (Fig. 2E). Correspondingly, ~70% of the 19*S* particles were detected in isolation ("free") (Fig. 2F). Overall, we found that 37% of synapses had one or more capped proteasome within a 1-μm radius, 35% of synapses had at least one free 20*S* particle within a 1-μm radius, and 78% of synapses had at least one free 19*S* particle within a 1-μm radius. None of the proteasome particles exhibited obvious clustering at synapses, consistent with previous observations (*6*). Although the free 20*S* proteasome has been implicated as an independent degradation machine with distinct protein substrates (*25*), the abundance of free 19*S* particles near synapses has not been previously detected.

We considered whether neurons have elevated 19*S* abundance near synapses to increase the likelihood of capping 20*S* particles (thus generating 26*S* and 30*S* proteasomes). To test this, we harvested somata- and neurite-enriched neuronal lysates from cortical neurons cultured in compartmentalized chambers (*9*) (Fig. 2G and fig. S1H). Using native protein gel electrophoresis, neuronal proteasomes comprising different assembly states (20*S*, 19*S*, 26*S*, and 30*S*) were separated as distinct bands that were visualized by immunoblotting against the 20*S* subunit PSMA7 or the 19*S* subunits PSMC1, PSMD6, and Uch37/Uchl5 (Fig. 2H) (*28*). In the neurite fraction, we found that ~50% of the total 20*S* immunofluorescence intensity was present in the free 20*S* band (Fig. 2I) and >60% of the total 19*S* immunofluorescence intensity was present in the free 19*S* band, consistent with our single-molecule localization data (Fig. 2, E and F). Approximately the same fraction of total 20*S* was present in the free 20*S* band from the somata-enriched fraction (Fig. 2I), but there was a significantly lower fraction of free 19*S* compared with the neurite-enriched fraction (Fig. 2J). These data indicate that the superstoichiometric presence of 19*S* in dendrites and synapses does not elevate the percentage of capped 20*S* proteasomal particles (26*S* or 30*S*), suggesting that the free 19*S* particles may perform a proteasome-independent function at synapses.

The free 19S is an independent Lys^{63} DUB

Although the 19*S* regulatory particle does not degrade protein substrates, it is known to cleave ubiquitin chains from ubiquitylated proteins at an early step in proteasome processivity (*29*). This 19*S* DUB activity is essential for subsequent protein degradation by the 20*S* particle and ubiquitin recycling. We thus hypothesized that the free 19*S* particles may operate as independent DUBs to reverse protein ubiquitylation, reasoning that this activity could serve to suppress proteasome degradation near synapses. We first examined the type of ubiquitin chain that is associated with the free 19*S* subunit (and the other proteasome assembly states) using native protein gel electrophoresis and antibodies that recognize either Lys^{48}-linked ubiquitin chains (Lys^{48}-ub) or Lys^{63}-linked ubiquitin chains (Lys^{63}-ub). As previously reported (*28*), we found that the Lys^{48}-ub label was mostly associated with 26*S* and 30*S* proteasomes (Fig. 3A) but not the free 19*S* particle.

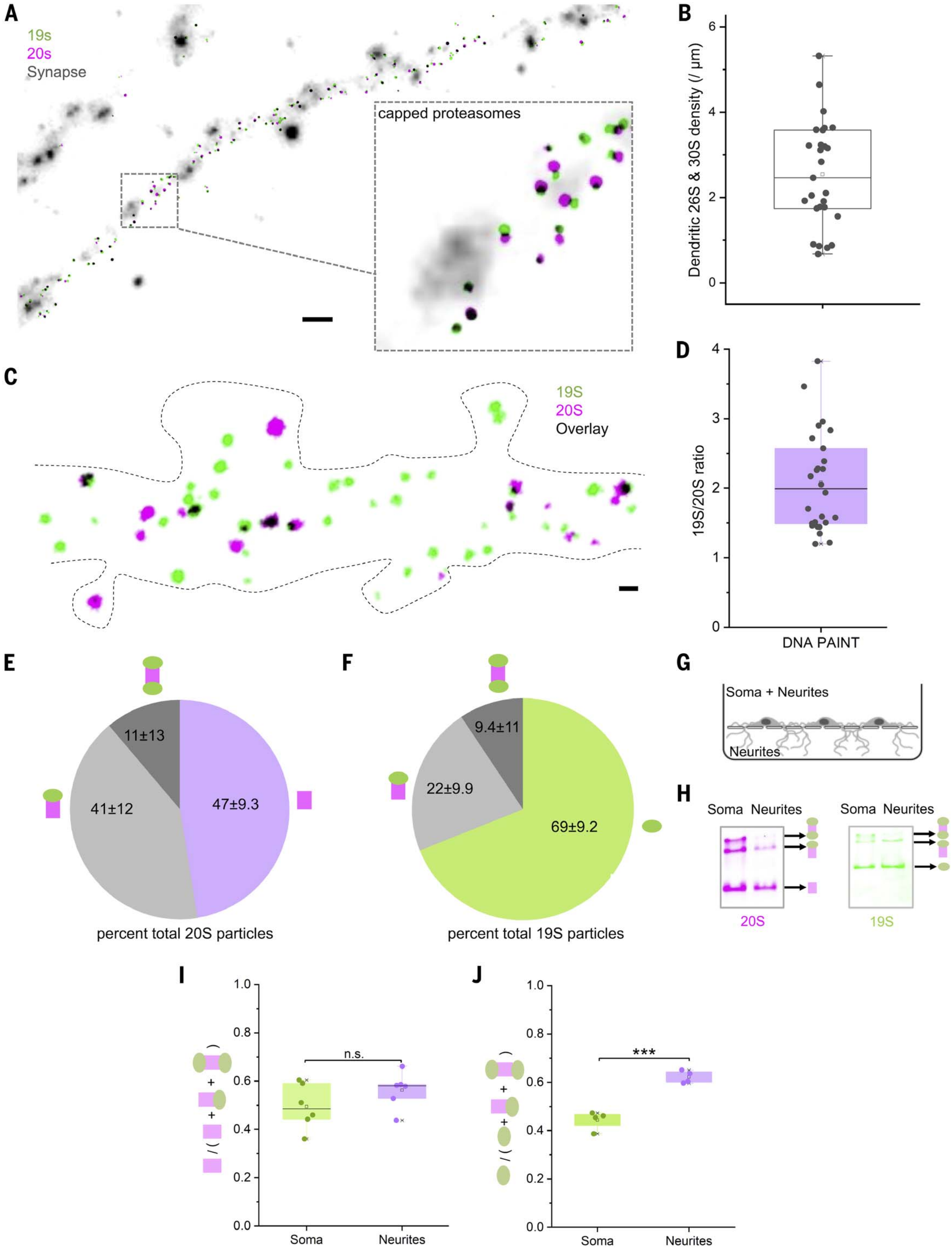


Fig. 2. Assembly states of 19*S* and 20*S* proteasomes in neuronal dendrites. (**A**) Single-molecule localization micrograph of super-resolved 26*S* and 30*S* proteasomes in neuronal dendrites and synapses (gray; immunostaining against Bassoon) from DNA PAINT, with 19*S* in green (anti-PSMC1/PSMD6), 20*S* in magenta (anti-PSMA7), and overlay in black. Scale bar, 2 μm. Inset shows an enlarged region containing multiple capped proteasomes as paired magenta and green puncta. (**B**) Scatterplot quantifying the abundance of 26*S* and 30*S* proteasomes along 27 dendrites from six neurons, with an average density of 2.5 ± 1.2 copies/μm of dendrite. (**C**) Dendritic segment with super-resolved 19*S* particle (green), 20*S* particle (magenta), as well as their colocalization in 26*S* and/or 30*S* proteasomes (in black). (**D**) Scatterplot showing the ratio of 19*S* particles versus 20*S* particles in 27 dendrites from six neurons obtained by DNA PAINT, with an average value >1 (1.9 ± 0.5). (**E**) Pie chart showing the percentage of dendritic 20*S* particles in solo, singly capped (26*S*), and doubly capped (30*S*) states. (**F**) Pie chart showing the percentage of dendritic 19*S* particles in free, singly capped (26*S*), and doubly capped (30*S*) states. The average values were obtained from 27 dendrites from six neurons. (**G**) Scheme of a compartmentalized chamber. (**H**) Native protein gel electrophoresis of soma-enriched lysate and neurite-enriched lysate immunoblotted against the 20*S* particle (magenta; anti-PSMA7) and 19*S* particle (green; anti-PSMC1/PSMD6/Uch37). The three bands in the anti-20*S* blot, starting from the top, are the 30*S*, 26*S*, and free 20*S* proteasomes, with the respective cartoons on the right. The three bands in the anti-19*S* blot, starting from the top, are the 30*S*, 26*S*, and free 19*S* proteasomes, with respective cartoons. (**I** and **J**) Box plots showing the fractions of free 20*S* (I) and free 19*S* (J) obtained from native protein gel electrophoresis of the soma-enriched and neurites-enriched lysates. Two-sample *t* tests detected no significant differences in the free 20*S* between the compartments and a significant increase in the free 19*S* in neurite-enriched lysates (***$P < 0.001$).

This indicates that the free 19*S* particle does not interact to an appreciable extent with Lys^{48}-ubiquitylated proteins in cell lysates (Fig. 3A). By contrast, we found that the Lys^{63}-ubiquitylated proteins were predominantly associated with the free 19*S* particle rather than with the 26*S* or 30*S* particles (Fig. 3B and fig. S2A).

To test directly whether Lys^{63}-ub chains undergo deubiquitylation as a result of their association with the free 19*S* particle, we reconstituted an enzymatic reaction that included purified free 19*S* particles and Lys^{63}-polyubiquitin chains in vitro. When the 19*S* particle was added to the purified Lys^{63}-ub chain, we observed the emergence of mono-ub products caused by

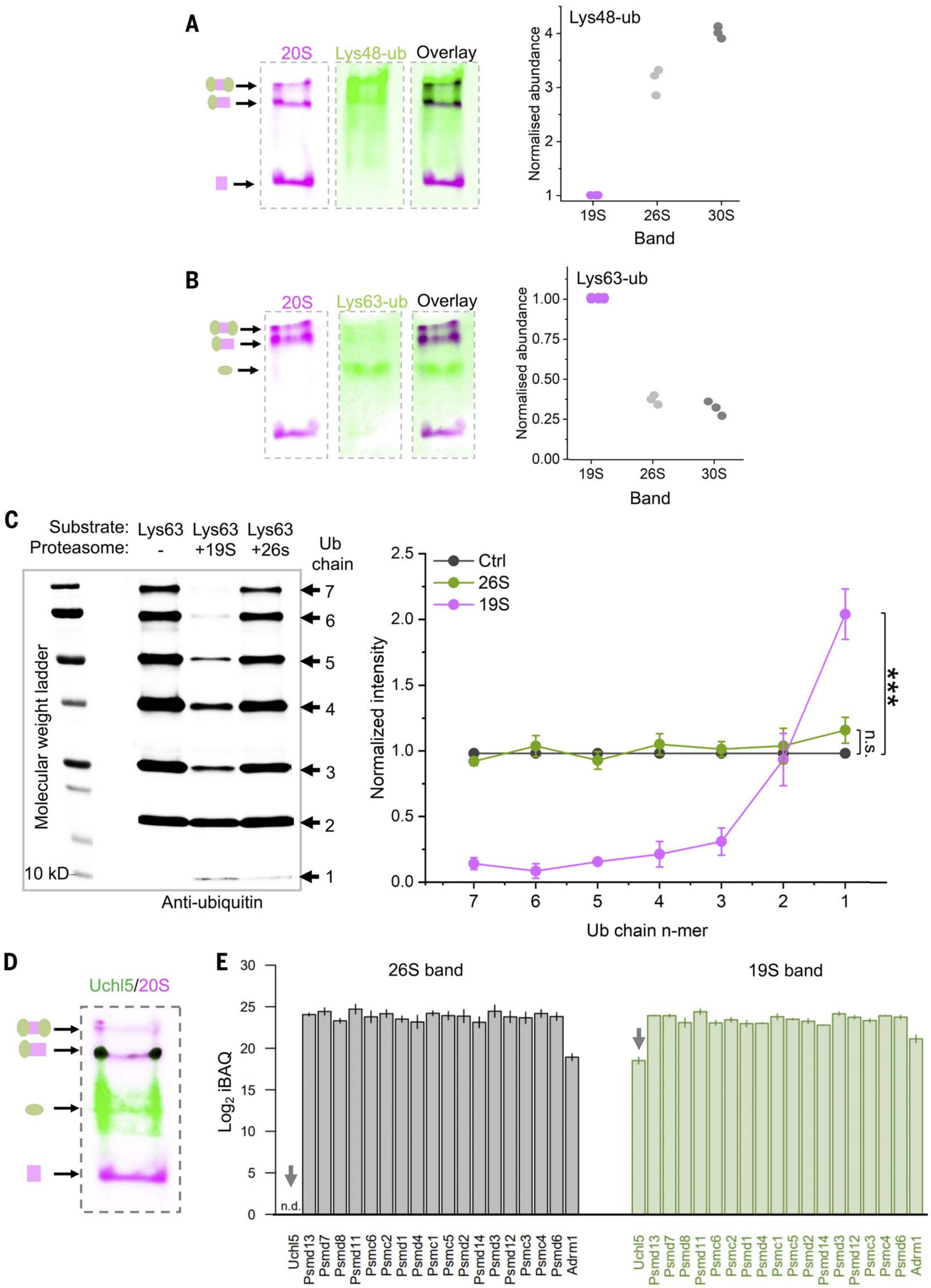


Fig. 3. Free 19*S* proteasome-regulatory particles function as proteasome-independent DUBs acting on Lys^{63}-ub. (**A** and **B**) Native protein gel electrophoresis showing the differential association of the 19*S* to linkage-specific ubiquitylation. (A) Predominant association of Lys^{48} ubiquitylation with 19*S* bound to 20*S* (26*S* and 30*S*; anti-PSMA7 and anti-Lys^{48}-ub). Cartoons on the left show 30*S*, 26*S*, and 20*S* from top to bottom. (B) Predominant association of Lys^{63} ubiquitylation with the free 19*S* band (anti-PSMA7 and anti-Lys^{63}-ub). Cartoons on the left show 30*S*, 26*S*, and 19*S* from top to bottom. Scatterplots showing the normalized ubiquitin intensities associated with free 19*S*, 26*S*, and 30*S* particles in native protein gel electrophoresis. (**C**) SDS-PAGE showing deubiquitylation of Lys^{63}-polyubiquitin chains by the free 19*S* particle after 2 hours of coincubation. Connected scatterplots show the normalized band intensities for different ubiquitin chain lengths in control (dark gray; normalized to 1), 19*S* (magenta), and 26*S* (green) lanes. Repeated-measures ANOVA detected a significant difference after coincubation with purified 19*S* particles (***$P < 0.001$) and no difference with purified 26*S* particles. (**D**) Native protein gel electrophoresis showing the predominant association of Uch37/Uchl5 with the free 19*S* particle. (**E**) MS analysis of excised 26*S* and 19*S* bands from native protein gel electrophoresis [e.g., in (D)] only detected Uch37/Uchl5 in the 19*S* band in addition to a variety of 19*S* subunits in both bands. Selected 19*S* protein (*x* axis) levels are shown by plotting the log2-transformed iBAQ protein intensities (*y* axis).

19*S* deubiquitylation, as well as the loss of higher-molecular-weight polyubiquitin species (Fig. 3C). To test whether this DUB activity is specific for the free 19*S* particle, we conducted the same experiment with the purified 26*S*. The 26*S* proteasome did not deubiquitylate Lys^{63}-ub chains to the same extent under the same conditions (Fig. 3C). This difference suggests that purified 19*S* may have additional DUBs that are absent or inactivated in the purified 26*S*. To investigate which 19*S* DUB(s) is responsible for this activity, we used specific antibodies to detect the association of the three different 19*S* DUBs, Rpn11/PSMD14, Uch37/Uchl5, and Usp14 (29), with either the 19*S* or the 26*S* proteasome. We found that purified 19*S* particles were associated with the DUBs Uch37/Uchl5 and Rpn11/PSMD14, whereas purified 26*S* particles were only associated with Rpn11/PSMD14 (fig. S2B). To determine whether Uch37/Uchl5 can deubiquitylate Lys^{63}-Ub, we inhibited its activity using b-AP15, a specific DUB inhibitor for Uch37/Uchl5 (and Usp14, which is absent in purified 19*S* and 26*S*; fig. S2B), but not PSMD14/Rpn11 (29, 30). Indeed, b-AP15 blocked the Lys^{63}-DUB activity of purified 19*S* in vitro (fig. S2C), strongly suggesting

that Uch37/Uchl5 is the DUB in the free 19*S* that deubiquitylates Lys^{63}-ub, rather than Rpn11/PSMD14. This Uch37/Uchl5 DUB activity is not observed in Uch37/Uchl5 that is not fully incorporated into 19*S* (*29*). Because the above purified 19*S* and 26*S* were not of neuronal origin, we prepared neuronal lysates, ran native gels, and then excised the 19*S* and 26*S* bands and determined the associated brain proteasomal proteins using mass spectrometry (MS). Although most proteasomal proteins were detected in both 19*S* and 26*S* particles, the neuronal Uch37/Uchl5 was predominantly associated with the free 19*S* in neuronal lysates (Fig. 3, D and E, and fig. S2, D and E). The preferential association of Uch37/Uchl5 (and Usp14; fig. S2E) with the free 19*S* may give rise to a distinct DUB activity, suggesting that free 19*S* exists as a separate pool from the DUBs that cap the 20*S* proteasome. Taken together, these data indicate that the free 19*S* particle can act as a proteasome-independent DUB against Lys^{63} ubiquitylation.

Synaptic proteins are major interactors of Lys^{63}-ub

Given the abundance of the free 19*S* near synapses, we considered whether synaptic proteins might be deubiquitylation targets. To identify the potential neuronal substrates of Lys^{63} ubiquitylation, we performed immunoprecipitation (IP) of Lys^{63}-ubiquitylated proteins using anti–Lys^{63}-ub tandem ubiquitin binding entities (TUBEs) in cortical neuron lysates (see the materials and methods) and subsequent MS-based proteomic analysis (figs. S3 and S4). We compared the Lys^{63}-ubiquitylated substrates with those identified by TUBEs-IP of Lys^{48} ubiquitylation and pan ubiquitylation (ub; with unspecified linkages). Among the >2000 proteins identified in the pull-down samples, 650 protein species in the Lys^{63}-ub IP were unique or enriched (compared with a "beads-only" pull-down background), whereas 430 or 1841 unique or enriched proteins were detected in the Lys^{48}-ub or pan-ub IP, respectively (Fig. 4A). Consistent with previous studies, several ribosomal proteins were Lys^{63} ubiquitylated (*29*). However, in addition, the Lys^{63}-ub IP contained many synaptic proteins (Fig. 4A), including essential presynaptic proteins (e.g., syntaxin, synaptogyrin, and synaptotagmin) and neurotransmitter receptors and ion channels [AMPA receptors (AMPARs), voltage-gated sodium channels, calcium channels, and potassium channels; see data S1]. By contrast, immunoprecipitation of Lys^{48}-ubiquitylated proteins mainly yielded hits that enriched for proteasomal components, consistent with its predominant association with the 26*S* and 30*S* proteasomes in native protein gel electrophoresis (Fig. 3A). The enrichment of synaptic proteins in the Lys^{63}-ub IP suggests that the free 19*S* may act as a DUB at synapses.

If the 19*S* DUBs regulate synaptic proteins, are there corresponding synaptic E3 ligases? Using synaptosome fractionation and MS, we detected 19 E3 ligases in the mouse cortical synaptosome fraction (see data S2) (*31*), including many well-studied Lys^{63}-ub ligases such as Nedd4, Nedd4l, and HUWE1 (*32*, *33*). To directly visualize some of the Lys^{63}-ub ligases at synapses, we conducted immunolabeling with antibodies against Nedd4, Nedd4I, and HUWE1 (fig. S5, magenta) in cultured rodent neurons and coimmunolabeled dendritic and synaptic protein markers (MAP2 and Bassoon; fig. S5, green). Confirming the MS detection in synaptosomes, we observed that all three E3 ligases were present in neuronal dendrites and dendritic spines but differentially distributed in some subcellular compartments. For example, consistent with its known role in axon development and repair (*34*), HUWE1 appeared abundant in axons (fig. S5C), whereas Nedd4 and Nedd4l were less abundant (fig. S5, A and B). Overall, these data indicate that neuronal synapses are equipped with the enzymes for the attachment and removal of Lys^{63}-ub.

The free 19S DUB regulates synaptic function

Among the many enriched synaptic targets for Lys^{63} ubiquitylation were proteins such as syntaxin, synaptotagmin, and synaptophysin, which are key players in synaptic vesicle release, suggesting a role for dynamic Lys^{63} ubiquitylation in the presynaptic regulation of synaptic transmission. To test whether the free 19*S* DUB regulates synaptic transmission, we again made use of b-AP15, the inhibitor of the 19*S* DUBs that are preferentially associated with the free 19*S* (Uch37/Uchl5 and Usp14) (Figs. 3, D and E, and fig. S2). We first confirmed that treatment with b-AP15 (5 μM for 1 hour) induced a significant accumulation of Lys^{63}-ubiquitylated proteins in cultured neurons (fig. S6, A and B), an effect that was not observed after treatment with conventional proteasome inhibitors that block either 20*S* protease activity (e.g., MG132) or proteasome-dependent deubiquitylation (e.g., capzimin, inhibitor of Rpn11/PSMD14) (fig. S6, A and B) (*35*). As expected, b-AP15 did not inhibit the proteolytic activity of neuronal proteasomes (fig. S6C), as measured by in-gel digestion of fluorogenic substrates (see the materials and methods). We then measured miniature excitatory postsynaptic currents (mEPSCs) in cultured rat cortical neurons (18 to 21 days in vitro) under control conditions and after treatment with b-AP15. Upon inhibition of 19*S* DUB activity with b-AP15, we observed a significant increase in mEPSC frequency (Fig. 4, B and C), suggesting that 19*S* DUB activity negatively regulates synaptic vesicle release. Consistent with this, immunoblotting showed that a Lys^{63}-ub–enriched presynaptic protein, synaptotagmin, showed increased ubiquitylation upon b-AP15 treatment (fig. S4E). Overall, these results indicate that free 19*S* DUB activity regulates synaptic vesicle release.

Within the Lys^{63}-ub interactome, we also detected postsynaptic targets including Gria1 [see data S1 and (*36*)], an essential subunit of the AMPAR complex that is responsible for most point-to-point excitatory synaptic transmission in the mammalian brain. Consistent with Gria1 being a target of 19*S* DUB activity, we found that the DUB inhibitor b-AP15 also induced a significant decrease in mEPSC amplitude (Fig. 4D) and a decrease in the initial slope of field excitatory postsynaptic potential (fEPSP) measured at the Schaffer collateral–CA1 synapses in rat hippocampal slices (Fig. 4E). To test directly whether the free 19*S* DUB activity regulates dendritic AMPARs, we visualized the total dendritic AMPAR population through immunofluorescence staining (using permeabilizing conditions) under control conditions or after b-AP15 incubation. Compared with control (Fig. 5A), we observed a significant loss in the total pool of dendritic AMPARs after just 1 hour of b-AP15 treatment (Fig. 5B and fig. S7A). Considering the ~4 day half-life of AMPARs (*24*), these data suggest that postsynaptic 19*S* DUB activity usually stabilizes the AMPAR pool, protecting it from degradation. This effect was not generalized to all transmembrane receptors, because the abundance of the transferrin receptor was not affected by b-AP15 (fig. S7C). Because AMPARs are degraded by lysosomes (*36*), we hypothesized that postsynaptic 19*S* DUB activity regulates the lysosomal degradation of AMPARs. To test this, we inhibited lysosomal function using chloroquine (40 μM, 1 hour) before b-AP15 incubation. Indeed, lysosome inhibition rescued the loss of dendritic AMPARs induced by b-AP15 (Fig. 5, A and B, and fig. S7A), suggesting that under normal circumstances the 19*S* DUB activity promotes AMPAR lysosomal degradation. By contrast, treatment with the proteasome inhibitor bortezomib or capzimin failed to prevent the b-AP15–induced loss (Fig. 5, A and B, and fig. S7B), indicating that the induced loss of AMPARs was proteasome independent. This latter observation is consistent with a previous study showing that proteasome activity did not alter AMPAR expression under basal conditions (*13*).

The above data suggest that the free 19*S* particles regulate the lysosomal degradation of AMPARs through Lys^{63} ubiquitylation, consistent with prior knowledge of Lys^{63}-ub signaling (*33*, *37*). To examine the importance of Lys^{63}-ub modifications on AMPAR levels, we expressed in cultured neurons a hemagluttinin (HA)–tagged ubiquitin mutant (K63R-HA) in which Lys^{63} in ubiquitin is replaced with an Arg, preventing Lys^{63}-ub chain formation. We compared the effects of K63R expression with overexpression of an HA-tagged wild-type ubiquitin. As expected, we observed reduced total

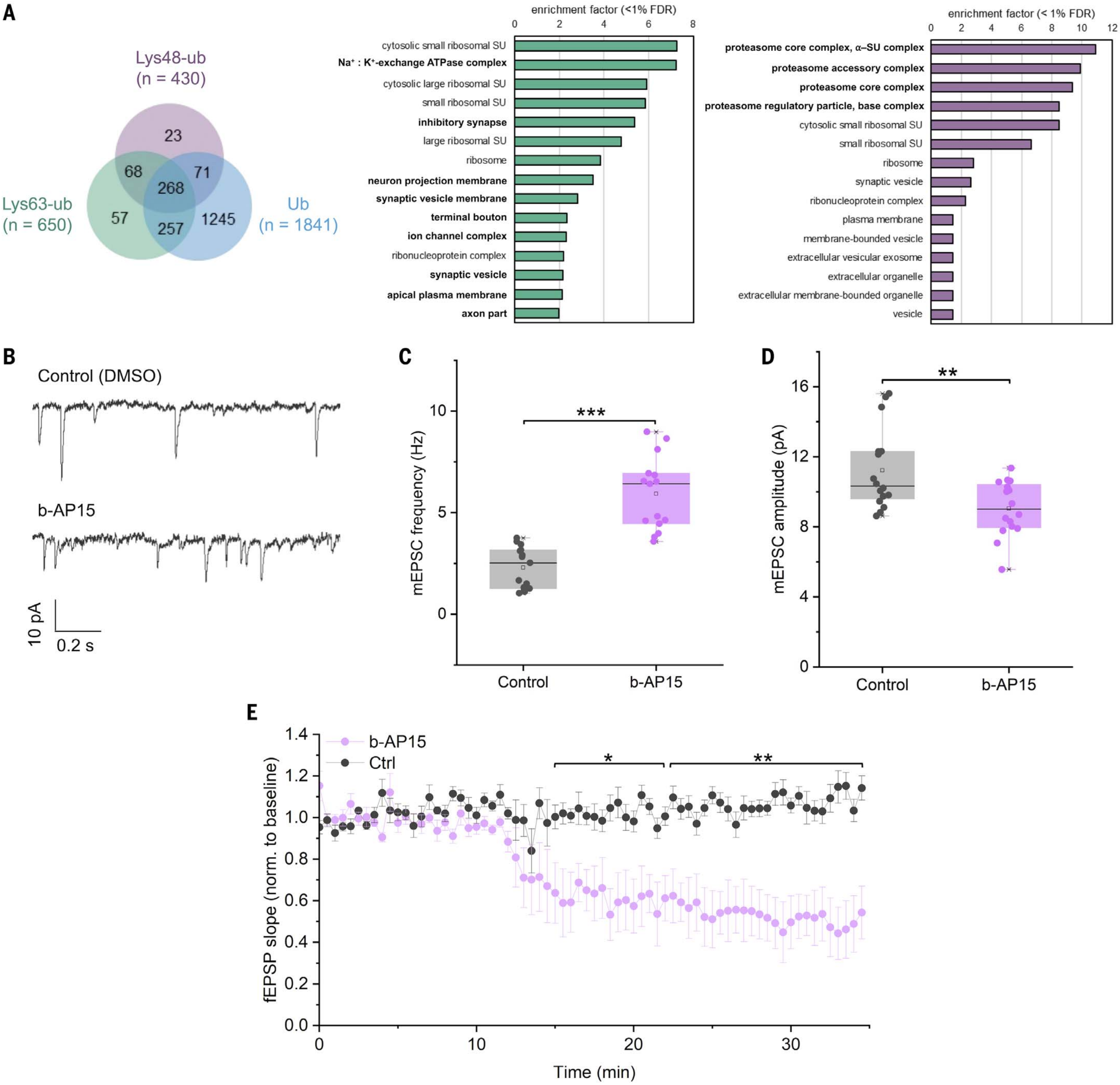


Fig. 4. Synaptic proteins are major interactors of Lys[63] ubiquitylation.
(**A**) Top 15 GO cellular component terms enriched in Lys[63]-ub (green; based on 650 of 2229 proteins) and Lys[48]-ub (purple; based on 430 of 2271 proteins) compared with total neuronal lysates. Proteins were quantified by MS and filtered based on uniqueness to the IP or fold enrichment (more than threefold) compared with an "empty-bead" control. GO term overrepresentation was analyzed compared with all proteins detected in the input samples (i.e., total neuronal lysates) using a Fisher's exact test with Benjamini-Hochberg correction (FDR <1%). (**B**) b-AP15, an inhibitor of 19*S* DUB activity, regulates miniature excitatory synaptic transmission, as shown by mEPSCs recorded from cultured rat cortical neurons under control conditions or with b-AP15 (5 μM, 1 hour) to inhibit 19*S* DUB activity. (**C**) Box plots showing the mEPSC frequency with and without b-AP15. (**D**) Box plots showing the mEPSC amplitude with and without b-AP15. Two-sample *t* tests detected a significant increase in mEPSC frequency and a significant decrease in mEPSC amplitude after b-AP15 treatment (***$P < 0.001$; **$P < 0.01$; 16 recordings from four biological replicates each). (**E**) Connected line plots showing electrophysiological recordings of evoked responses (fEPSP slope normalized to baseline; *y* axis) in acute rat hippocampal slices with b-AP15 (magenta) or with DMSO as a control (dark gray) over time (*x* axis). Two-sample *t* tests detected a significant decrease in the initial slope of the fEPSP response after b-AP15 application (*$P < 0.05$, **$P < 0.01$; five biological replicates each).

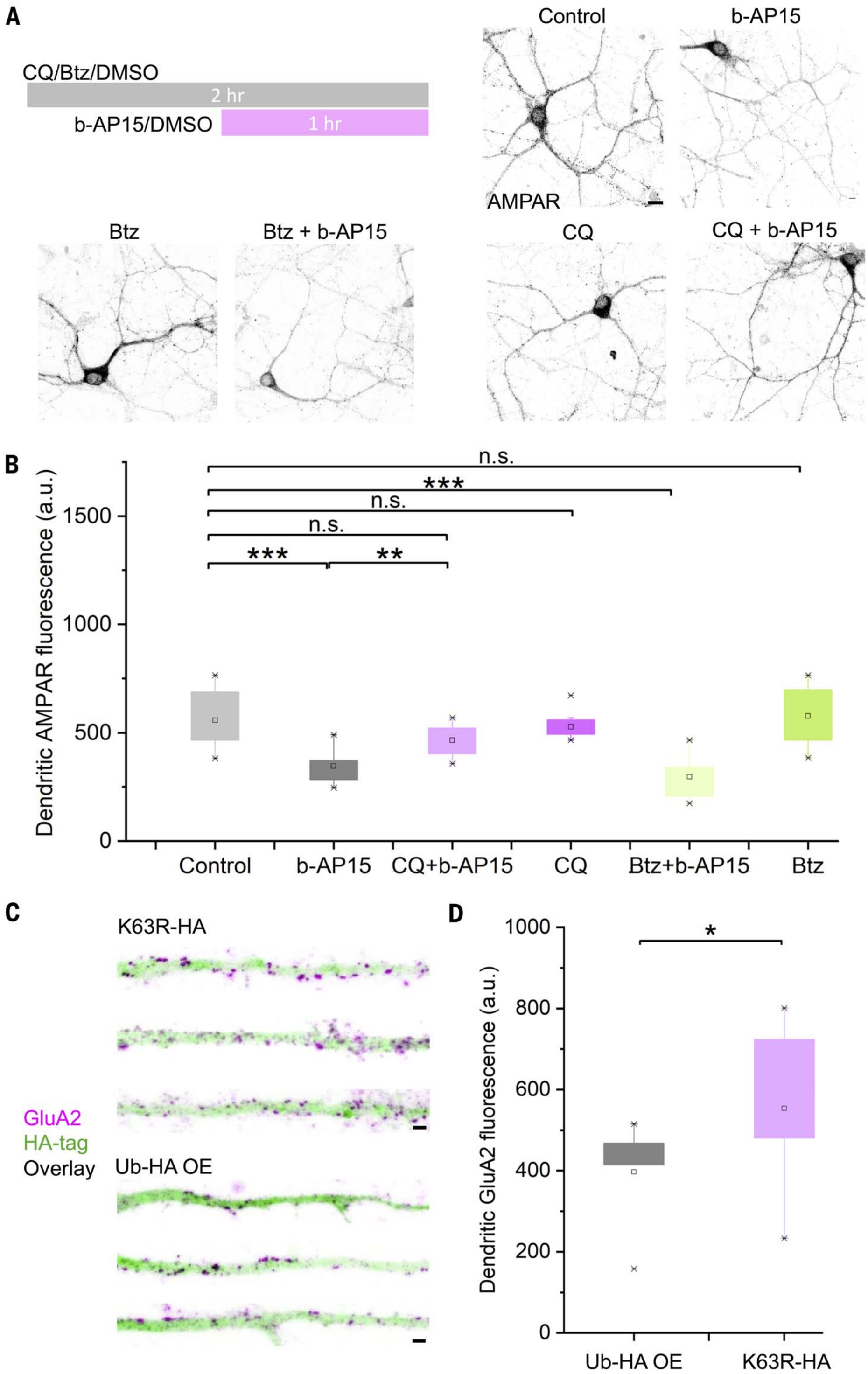


Fig. 5. Inhibition of (proteasome-independent) 19*S* DUB activity regulates dendritic AMPAR abundance through Lys^{63} ubiquitylation.
(**A**) b-AP15, a 19*S* DUB inhibitor, induced a loss of dendritic AMPARs that was lysosome dependent and proteasome independent. Top left scheme shows the time course of the pharmacological treatment, in which cultured rat cortical neurons were treated with a lysosome inhibitor (chloroquine, CQ), a proteasome inhibitor (bortezomib, Btz), or DMSO as a control for 1 hour before the addition of b-AP15 (or DMSO for the control) for another 1 hour. The corresponding confocal micrographs immunolabeled for the total AMPAR population (anti-GluA2) were labeled after each treatment. Scale bar, 10 μm. (**B**) Box plots showing the distribution of total dendritic AMPAR fluorescence intensity (in arbitrary units, a.u.) for each treatment shown in (A). Each treatment group contains >50 dendrites from >10 neurons. ANOVA with post hoc Bonferroni test indicated significant differences between individual neurons in control and b-AP15, between control and CQ and b-AP15, and between b-AP15 and CQ and b-AP15 (***P* < 0.01, ****P* < 0.001). (**C**) Confocal micrographs of straightened dendrites from transfected cultured neurons expressing K63R-HA or overexpressing (OE) ub-HA. Samples were immunolabeled for the HA-tag (green) to identify transfected neurons and for GluA2 (magenta) to visualize AMPAR abundance. (**D**) Box plots showing the total dendritic AMPAR fluorescence intensity (a.u.) in transfected neurons that either overexpressed wild-type ubiquitin or expressed K63R ubiquitin. A two-sample *t* test detected a significant increase in the total dendritic AMPAR population in individual neurons of the K63R group (**P* < 0.05; >30 dendrites from nine transfected neurons in each group).

Lys^{63} ubiquitylation in neurons that overexpressed the dominant-negative K63R ubiquitin (fig. S8). We also detected elevated total AMPAR levels in neurons expressing K63R relative to control neurons, which overexpressed wild-type ubiquitin (ub-HA) (Fig. 5, C and D, and fig. S8A). Together with the bAP15-induced reduction in total AMPAR levels, these data indicate that Lys^{63} ubiquitylation works to destabilize total AMPAR levels, whereas the 19*S* DUB works to stabilize them.

Our above-described in vitro assays (Fig. 3C) identified Uch37/Uchl5 as a free 19*S* DUB with activity against Lys^{63}-ub and a specific target of b-AP15. To directly test whether Uch37/Uchl5 regulates AMPARs, we knocked down Uch37/Uchl5 using a short hairpin RNA (shRNA) strategy (see the materials and methods). Compared with control (transfected with a scrambled sequence), we detected a ~60% knockdown (KD) of Uch37/Uchl5 by immunofluorescence staining (fig. S9, A and B). Neurons with Uch37/Uchl5 KD showed a significant decrease in dendritic AMPAR abundance compared with control neurons (Fig. 6). KD of the other 19*S*-associated DUB, Usp14, also reduced dendritic AMPAR abundance, although Usp14 showed

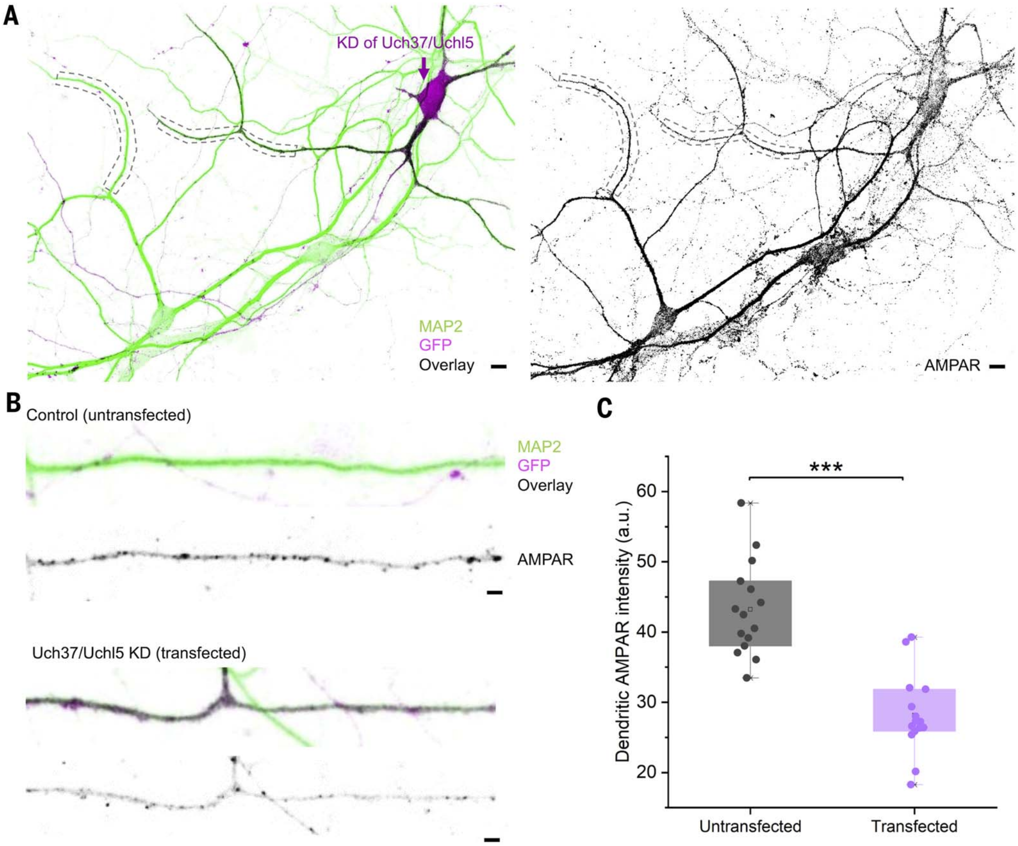


Fig. 6. KD of a 19*S* DUB regulates dendritic AMPAR abundance. (**A**) KD of a 19*S* DUB, Uch37/Uchl5, induced a loss of AMPARs. Left confocal micrograph shows a cultured neuron transfected to express green fluorescent protein (magenta) and shRNAs against Uch37/Uchl5. All neurons were immunostained using MAP2 antibodies (green). Right confocal micrograph shows AMPAR immunofluorescence (anti-GluA2). Scale bar, 5 μm. (**B**) Example dendritic branches [indicated by dashed boxes in (A)] showing a decrease in dendritic AMPAR abundance upon Uch37/Uchl5 KD. Scale bars, 2 μm. (**C**) Box plots showing dendritic AMPAR abundance in control (untransfected) neurons and Uch37/Uchl5-KD neurons (transfected). A two-sample *t* test found a significant decrease in individual Uch37/Uchl5-KD neurons (***$P < 0.001$; 15 neurons each).

mostly nuclear localization (fig. S9, C to E). Thus, consistently, conditions that stabilize AMPAR Lys^{63}-ub linkages (e.g., b-AP15 treatment or 19*S* DUB KD) and those that destabilize them (K63R) had opposite effects on the total AMPAR pool. This directly links Lys^{63} ubiquitylation to AMPAR abundance, and, together with the above data, suggests that Lys^{63} ubiquitylation normally destabilizes the AMPAR pool, whereas the 19*S* activity opposes this destabilization.

In the above experiments, inhibition of 19*S* DUB activity (by b-AP15) induced a loss of AMPARs, which presumably involves accelerated endocytic sorting, a step required for surface AMPARs to enter lysosomal degradation (*33*). To directly test whether loss of 19*S* DUB function accelerates AMPAR endocytosis, we performed live-cell surface labeling of AMPARs followed by specific immunolabeling of endocytosed surface AMPARs after 1 hour of internalization, as described previously (*13*). Unexpectedly, we found that AMPAR internalization was reduced, rather than increased, upon b-AP15 treatment (Fig. 7, A to C, and fig. S10A). It thus appears that the 19*S* DUB activity normally facilitates AMPAR internalization, in contrast to previously known DUBs (e.g., Usp8) (*33*). However, like other DUBs (*33*), 19*S* DUB also stabilizes the total (internal and surface) AMPAR pool (Fig. 5).

At first glance, the above results appeared inconsistent. We considered the possibility that differences in the number or distribution of surface AMPARs might provide an explanation. To directly test how 19*S* DUB activity affects surface AMPARs, we performed surface immunolabeling of AMPARs with and without the 19*S* DUB inhibitor b-AP15. In contrast to the b-AP15–induced reduction in total (internal and surface) dendritic AMPARs, the overall abundance of the surface AMPARs was indeed unchanged (Fig. 7, D to F, and fig. S10B). Instead, the distribution of the surface AMPARs appeared altered: The receptors were less clustered and punctate after b-AP15 treatment (Fig. 7, E and G). Previous work has shown that AMPARs are mobile in the neuronal membrane but exhibit increased dwell times at synapses, resulting in their clustering (*38*). To assess whether Lys^{63}-ub regulates AMPAR movement, we measured the mobility of surface AMPARs through single-particle tracking in live neurons under control conditions and after treatment with b-AP15 (Fig. 7H). Inhibition of 19*S* DUB activity significantly elevated the mobile fraction of surface AMPARs (maximum velocity is shown in Fig. 7I and mean velocity in fig. S10C), resulting in reduced clustering (Fig. 7G). In the maximum velocity analysis, we observed two major populations of AMPARs in both control and b-AP15 groups (Fig. 7I), a population that was largely immobile and a population that was (momentarily) mobile. Although the velocities of these two populations appeared to be unchanged by b-AP15, the fraction of mobile AMPARs increased with b-AP15 (Fig. 7I). These data suggest that Lys^{63} ubiquitylation of AMPARs promotes receptor mobility that can be negatively regulated (stabilized) by the free 19*S* DUB activity. Treatment with b-AP15 blocks this stabilization and results in reduced AMPAR clustering at synapses, impairing postsynaptic sensitivity to spontaneous presynaptic vesicle fusion (reduced mEPSC amplitudes; Fig. 4D). Because surface AMPARs are endocytosed at sites distant from synapses (*39*), it thus appears that the free 19*S* subcomplex stabilizes the (presumably synaptic) clustering of surface AMPARs and yet facilitates their internalization once they diffuse away. Once internalized, the free 19*S* subcomplex then stabilizes the reserve pool of AMPARs, preventing them from degradation. In addition, we found that free 19*S* facilitated AMPAR exocytosis, providing an alternative to their degradation fate (fig. S10D). These results demonstrate that proteasome-independent 19*S* DUB activity regulates AMPARs at multiple trafficking points, and all of these actions work together to maximize receptor availability at synapses (Fig. 7J).

Discussion

Up to 90% of the total proteome of a single neuron resides in dendrites and axons (*1*, *3*), indicating that the relative protein turnover

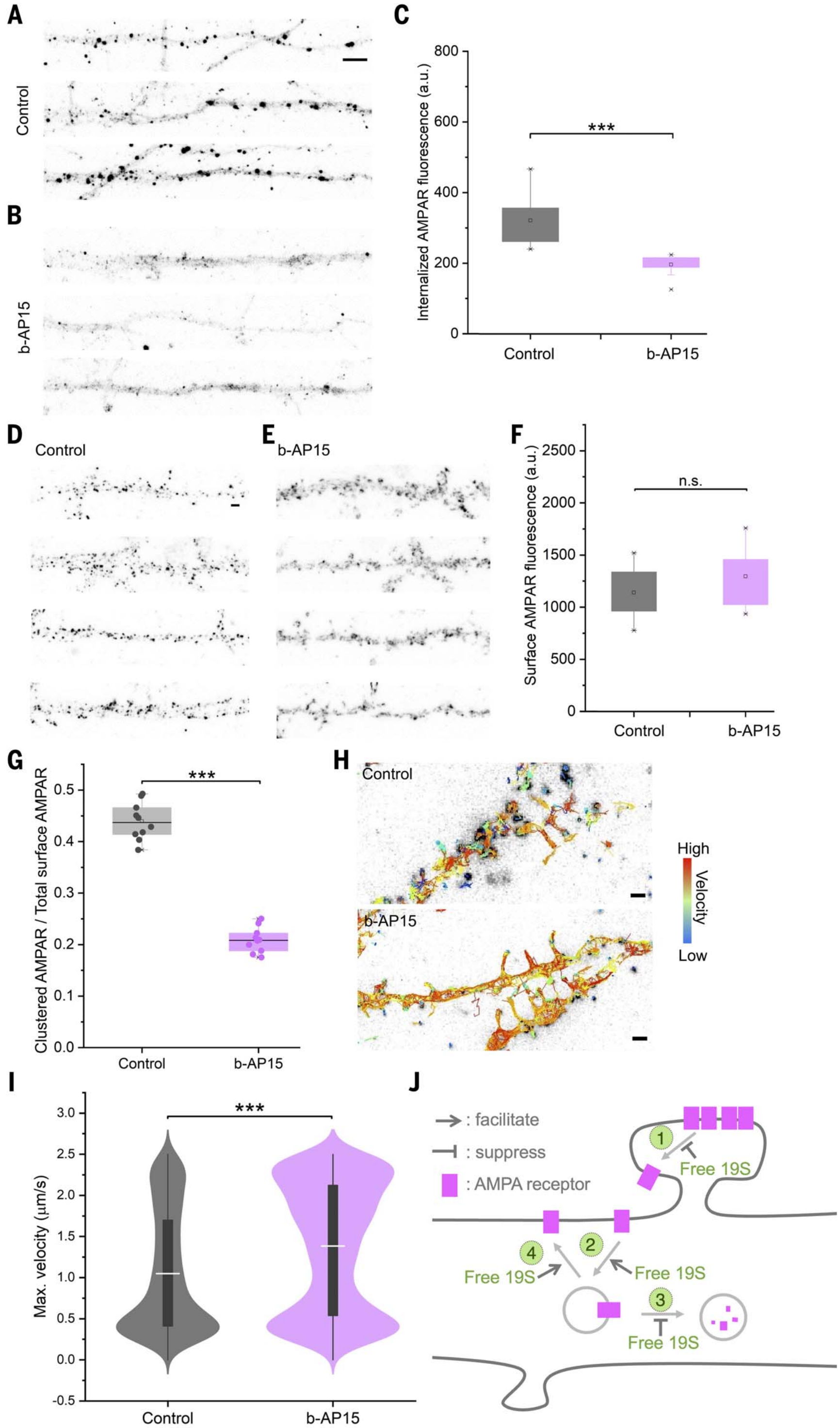


Fig. 7. Postsynaptic 19*S* DUB activity maximizes the synaptic availability of AMPARs. (**A** and **B**) Confocal micrographs showing internalized AMPARs in cultured rat cortical neuronal dendrites without (A) and with (B) b-AP15. Scale bar, 2 μm. (**C**) Box plots showing the distribution of internalized AMPAR fluorescence intensity with and without b-AP15 treatment. A two-sample *t* test found a significant decrease in individual neurons after b-AP15 treatment (****P* < 0.001; >50 dendrites from >10 cells each group). (**D** and **E**) Confocal micrographs showing surface AMPARs on straightened dendrites without (D) or with (E) b-AP15 treatment. Scale bar, 2 μm. (**F**) Box plots showing the level of surface AMPAR fluorescence in straightened neuronal dendrites with (63 dendrites from 10 neurons) or without (48 dendrites from 10 neurons) b-AP15 treatment. A two-sample *t* test indicated no significant difference between individual neurons of the two groups. (**G**) Box plots showing the fraction of clustered AMPARs on cell surface with or without b-AP15 treatment. A two-sample *t* test detected a significant decrease in individual neurons with b-AP15 treatment (****P* < 0.001; >10000 clusters from >10 cells in each group). (**H**) Single-particle tracking of surface AMPARs with tracking traces overlaying on the first frame. Color represents the average velocity in each tracking trace (warmer colors indicate higher velocity). There are 2636 tracks for control and 3397 tracks for b-AP15. Scale bar, 1 μm. (**I**) Violin plots showing the distributions of surface AMPAR mobility (maximum velocity within the tracking duration) with and without b-AP15 treatment. A two-sample *t* test indicated a significant increase with b-AP15 (****P* < 0.001; 2636 tracks for control and 3397 tracks for b-AP15). (**J**) Schematic showing that free neuronal 19*S* DUB activity regulates the synaptic availability of AMPARs at multiple trafficking steps including (1) increasing the synaptic pool, (2) favoring the endocytosis of nonsynaptic receptors, (3) reducing lysosomal degradation, and (4) facilitating exocytosis in a concerted, proteasome-independent manner.

demand at synapses is huge (40, 41). Local proteasomal function is thus essential for synaptic homeostasis as well as plasticity. Here, we revealed a surprising differential distribution of the proteasomal catalytic and regulatory particles in neuronal cell bodies, dendrites, and axons. We quantified the fractions of proteasomal particles in different proteasomal assembly states and demonstrated the presence of excess, free 19*S* regulatory particles near synapses. We discovered that these free 19*S* particles deubiquitylate important synaptic proteins that are Lys^{63} ubiquitylated, such as AMPARs. The dynamics of Lys^{63} ubiquitylation and deubiquitylation potently regulate synaptic transmission in both the pre- and postsynaptic compartments. The excess of the free 19*S* particles present in dendrites thus represents an interesting example in which a subcomplex of a large protein machine performs a moonlighting function near synapses, in this case, as an independent "editor" of a synaptic ubiquitin code.

We showed that neuronal subcellular compartments differentially localize the two main (19*S* regulatory and 20*S* catalytic) proteasomal particles. This challenges the notion that

biological machines exist as a single entity throughout their life cycles. Instead, it appears that neurons have taken advantage of distinct functions of 19*S* particles and adapted their use to fine-tune synaptic strength. Previous studies have shown that oxidative stress leads to 26*S* proteasome decapping (*42*), suggesting a potential source of free 19*S*. However, it is also possible that the free 19*S* may exist as a separate pool from the 19*S* that is assembled with the 20*S*. What constitutes the minimal 19*S* subcomplex necessary for DUB activity? It is known that the 19*S*-associated DUBs (Uch37/Uchl5 and Usp14) are autoinhibited when expressed alone (*29*), but how many 19*S* protein components are needed to assemble a free 19*S* DUB? In our experiments the "free" 19*S* detected by microscopy and native gels could represent a mixture of 19*S* assembly states that are distinct from the 20*S*-bound 19*S* in both conformation and subunit composition. The possibility of different 19*S* formats would require different mechanisms for neuronal proteasome biogenesis. Previous studies have shown that neuronal proteasomal transcripts are predominantly somatic (*43*), suggesting that proteasomal subunits are synthesized and assembled in the cell body (*16*, *44*). However, the transcripts of proteasomal subunits are not stoichiometric (*45*). To maintain stoichiometric expression, quality control pathways engage to rapidly degrade excess proteasome subunits, as well as their partial assembly states that are "orphaned" from 19*S* or 20*S* assembly (*45*). To survive the surveillance of quality control, proteasome subunits presumably must assemble into 19*S* and 20*S* particles rapidly after translation. Once assembled, however, it remains unknown how 19*S* and 20*S* are sorted into different neuronal subcellular compartments (*46*).

We identified the free 19*S* regulatory particle as a distinct editor of synaptic ubiquitin. For example, inhibition of proteasome-independent 19*S* DUB activity led to the accumulation of Lys^{63} ubiquitylation and resulted in abnormal presynaptic neurotransmitter release. In parallel, we identified multiple presynaptic proteins involved in synaptic vesicle exocytosis as substrates and/or interactors of Lys^{63} ubiquitylation. For example, multiple isoforms of syntaxin (stx1a, stx1b, stx7, and stx12), a family of membrane-anchored SNARE proteins essential for membrane fusion, were identified in our data. Syntaxin has consistently been shown to have Lys^{63}-ub–binding domains (*47*). Other important classes of synaptic proteins were also identified in the Lys^{63}-ub interactome as potential targets regulated by free 19*S*, including synaptic adhesion molecules (neuroligins), ion channels, and neurotransmitter receptors. Finally, the DUB activity associated with the free 19*S* near synapses may also facilitate ubiquitin recycling, because its inhibition led to the accumulation of polyubiquitin chains. Previous studies have shown that accumulation of Lys^{48}-ubiquitylated proteins upon proteasome inhibition also leads to altered synaptic transmission (*48*). Therefore, the free 19*S* synaptic DUB works in concert with the diverse local E3 ligases that we and others have identified. Both 19*S*-associated DUBs (Usp14 and Uchl5) have been implicated in neurological disorders such as ataxia and abnormal brain development (*49*, *50*). So far, hundreds of E3 ligases and DUBs have been identified in mammalian cells as emerging drug targets (*15*, *51*, *52*), with many present at synapses (*31*, *33*, *53*, *54*).

We discovered one important function of the free 19*S* DUB activity: the concerted regulation of many aspects of the AMPAR life cycle, all of which conspire to maximize the synaptic availability of AMPARs. The free 19*S* proteasome directs the flow of AMPARs toward synapses at four main trafficking points: (i) at the plasma membrane, it facilitates AMPAR clustering, which is evident by the loss of clustering upon inhibition of 19*S*; (ii) in the extrasynaptic space, it facilitates AMPAR endocytosis; (iii) once AMPARs are internalized, the free 19*S* suppresses their degradation; and (iv) the free 19*S* also facilitates the exocytosis of AMPARs. It has been long known that 19*S* is sequestered by synapses during elevated synaptic activity (*6*). However, until now, this synaptic localization has only been associated with elevated local protein degradation. Our work provides an additional dimension to the role of proteasome particles in synaptic regulation: The functional states of a key synaptic protein can be regulated by free 19*S* through the regulation of an alternative ubiquitin chain.

Materials and methods

Cell culture and antibody staining

Dissociated rat primary cortical neuron cultures were prepared and maintained as previously reported for hippocampus neurons (*5*). In brief, cortices from postnatal day 1 rat pups of either sex (RRID:RGD_734476; strain Sprague-Dawley) were dissected and dissociated by incubating with L-cysteine-papain solution at 37°C before being plated onto MatTek dishes (MatTek, Ashland, MA) previously coated with poly-D-lysine for microscopy or onto 10-cm petri dishes (MatTek, Ashland, MA) for biochemistry. Cultured cells were incubated in Neurobasal A medium (Invitrogen, Carlsbad, CA) supplemented with B-27 (Invitrogen) and Glutamax (Invitrogen) at 37°C and 5% CO_2 for 18 to 21 days before use. Neuron cultures in compartmentalized chambers were prepared and maintained as described previously (*9*). Cultured neurons can extend their neurites (dendrites and axons) through the porous membranes into the bottom compartment while their cell bodies (and some processes) remain in the upper compartment. All procedures followed national animal care guidelines and the guidelines issued by the Max Planck Society and were approved by local authorities.

Immunolabeling and microscopy

Cultured cortical neurons in MatTek dishes were fixed in paraformaldehyde-sucrose [4% paraformaldehyde; Alfa Aeser, 4% sucrose in phosphate-buffered saline containing $MgCl_2$ and $CaCl_2$ (PBS-MC)] at room temperature (RT) for 20 min, washed three times in PBS before permeabilization with 0.5% Triton X-100 in 1× PBS, pH 7.4, for 15 min and blocking in PBS containing 4% goat serum (Gibco) for 1 hour.

For immunofluorescence staining and visualization by confocal microscopy, primary and secondary antibodies are described in Table 1.

Unless otherwise noted, overnight primary antibody staining (1:1000) was performed in PBS containing 4% goat serum (Gibco) followed by three washes with PBS for 5 min each. Secondary antibody staining (1:1000) was performed for 1 hour in PBS, followed by three washes with PBS for 5 min each. A Zeiss LSM 880 confocal setup was used for image acquisition with airy-scanning capacity (lateral resolution of ~120 nm). All antibodies were validated by Western blot and/or immunofluorescence staining before use (e.g., fig. S1).

For single-molecule localization microscopy, secondary antibody staining of proteasome particles for DNA PAINT used antimouse antibodies conjugated to a single-stranded DNA oligo (P1 docking) and anti-rabbit antibodies conjugated to a P5-docking oligo (1:1000, custom made, sequences as previously reported) (*27*). After washing three times in PBS (5 min each), neurons were briefly fixed (5 min) and then stored in PBS at 4°C for up to 3 weeks until DNA-PAINT imaging commenced. For DNA-PAINT imaging, an imaging buffer containing 500 pM P1 or P5 imager oligo conjugated with Atto655 (Eurofins Genomics) in 500 mM NaCl in PBS, pH 7.4, was used as previously described (*4*). Gold fiducial markers (90 nm, A1190, Nanoparz) were sparsely plated. PAINT was performed on a Leica DMi8 S system with infinity TIRF HP, Infinity Scanner, and an iXon Ultra 888 EMCCD camera (Andor). A 100× oil-immersion objective (HC PL APO CORR TIRF, NA 1.47) was used in combination with a motorized TIRF illuminator. Image acquisition was performed with 638-nm excitation wavelength (150 mW) and the LAS X software package. To exchange imager oligos, samples were washed three times with 500 mM NaCl in PBS before a new imaging buffer containing an alternative imager oligo was added.

Wide-field micrographs of Bassoon and MAP2 reference markers (see above description of immunofluorescence labeling) were acquired for distal dendrites (at least one branch point away from the cell body) before DNA-PAINT.

Table 1. Primary and secondary antibodies used for immunofluorescence staining and visualization.

Antibody	Source	Identifier
Guinea pig anti-MAP2	Synaptic Systems	188004
Guinea pig anti-Bassoon	Synaptic Systems	141004
Mouse anti-alpha7/PSMA7	Enzo Life Sciences	MCP72 clone; PW8110
Mouse anti-alpha 1,2,3,5,6,7	Enzo Life Sciences	PW8195
Mouse anti-Rpt5/PSMC3	Enzo Life Sciences	TBP1-19 clone; PW8770
Rabbit anti-Rpt2/PMSC1	Thermo Fisher	HPA000872
Rabbit anti-Rpn7/PSMD6	Abcam	ab247004
Rabbit anti-ub	Cell Signaling	43124
Rabbit anti–Lys^{48}-ub	Cell Signaling	4289
Rabbit anti–Lys^{63}-ub	Cell Signaling	5621
Rabbit anti-Uchl37/Uchl5	Abcam	ab133508
Rabbit anti-Usp14	Abcam	ab137432
Rabbit anti-synaptotagmin 1,2	Abcam	ab131548
Mouse anti-HA	Roche	10952100
Rabbit anit-Nedd4l	Cell Signaling	4013
Rabbit anti-Nedd4	Cell Signaling	5344
Rabbit anti-HUWE1	Abcam	ab271032
Mouse anti-GluA2	Gift from Gouaux lab, OHSU	N/A
Chicken anti-HA	Abcam	ab9111
Donkey anti-mouse Cy3	Jackson ImmunoResearch	715-165-150
Goat anti-mouse IgG1 Alexa Fluor 488	Invitrogen	A21121
Goat anti-rabbit Alexa Fluor 647	Invitrogen	A32733
Goat anti-mouse IgG2a Alexa Fluor 647	Invitrogen	A21241
Goat anti-guinea pig Alexa Fluor 488	Invitrogen	A11073
F(AB)2 anti-mouse QDOT 565	Thermo Fisher	Q11032MP

Super-resolution acquisition was performed with HILO illumination and a power of 30 to 40 mW (determined directly after the objective and under wide-field configuration). Time-lapse datasets with 50,000 frames and 16-bit depths were acquired for each proteasome particle localization with 5-Hz frame rate and 10-MHz camera readout bandwidth. Gain list was 2, and the electron-multiplying gain was 100.

DNA-PAINT data were processed as previously reported (*5*). Briefly, the acquisitions were reconstructed with Picasso:Localize, a module of the Picasso software, by applying a minimal net gradient of 15,000. Drift corrections were applied with Picasso:Render. Drift-corrected data were filtered using Picasso:Filter. Raw localizations within a maximal distance of 6× measured localization precision and showing a maximum number of transient dark frames of 20 were linked together, resulting in a single, linked localization event. As shown previously (*5*), we observed minimal bleaching during imaging. In the number-of-linked-localization distribution for all the identified clusters, the smallest population of clusters was attributed as the signal of a single protein copy, with a localization precision (measured by the nearest-neighbor analysis) of 13.1 nm. To identify different assembly states of proteasomes, coincidence detection thresholds for the interdistance between proteasome 19*S* and 20*S* clusters were set at 100 nm, taking into consideration of the localization precision, oligo-docking site interdistance, image drift correction, and the actual distance between the 19*S* and 20*S* epitopes. The resulting fractions of proteasome particles in each assembly state were verified by orthogonal methods including biochemistry (Fig. 2) and existing cryo–electron tomography data (*23*). Because of the technical limitations associated with the method, there was no three-dimensional information in our DNA PAINT, leaving the possibility for a 30*S* to be mistaken as 26*S*. However, a comparison with the orthogonal methods mentioned above suggests that such misidentifications were likely minimal.

Labeling of surface AMPARs (including for confocal microscopy and single-particle tracking) and internalized and exocytosed AMPRs was conducted as previously described (*13*, *55*). Briefly, cultured rat cortical neurons were incubated with anti-GluA2 antibodies (1:100) at 37°C with 5% CO_2 for 5 min, followed by washing with conditioned medium to remove excess free antibodies. For surface AMPAR confocal microscopy, cultured neurons were then fixed without permeabilization and stained with fluorescently tagged secondary antibodies before imaging. For single-particle tracking, living primary antibody–labeled cultured neurons were then live labeled with secondary antibodies tagged by quantum dots for 5 min before washing and subsequent live-cell imaging (frame rate: 5 Hz; duration: 15 min). Single-particle tracking data were analyzed by the plugin TrackMate in Fiji. For visualization of internalized AMPARs, cultured neurons were incubated with anti-GluA2 antibodies for 20 min at 37°C with 5% CO_2. The samples were then washed once with cold PBS-MC, incubated on ice with cold 0.5 M NaCl and 0.2% acetic acid for 4 min to strip extracellular receptor-bound antibodies, and then washed again with cold PBC-MC. The samples were then fixed, permeabilized, and stained with fluorescent secondary antibodies for imaging. As a control, neurons that were stained without permeabilization showed negligible signal. For visualization of exocytosed AMPARs, surface AMPARs of live cultured neurons were blocked by anti-GluA2 antibodies and an unlabeled secondary antibody for 5 min at 37°C with 5% CO_2. The excess antibodies were washed away, and the cultures were allowed 1 hour in the incubator for exocytosis. Then, the neurons were fixed (without permeabilization), and the newly exocytosed AMPARs were labeled with anti-GluA2 antibodies and a fluorescent secondary antibody (gt anti-ms; Alex Fluor 488) for microscopy.

Biochemistry

For native protein gel electrophoresis, cortical neurons cultured in 10-cm petri dishes were washed in ice-cold PBS buffer before they were scraped and lysed on ice using a native lysis buffer containing 50 mM Tris, pH 7.4, 1 mM dithiothreitol, 5 mM $MgCl_2$, 2 mM ATP, and 250 mM sucrose. Homogenization was performed using a 1-ml syringe with a 27-guage needle before clarification at 4°C at 13 × 10^3 relative centrifugal force (rcf) for 5 min. The clarified lysate was supplemented with native sample buffer (1610738, Bio-Rad) and used for native protein gel electrophoresis. The running buffer for native protein gel electrophoresis contained 500 ml of Tris-borate buffer (T4415, Merck), 276 mg of ATP dissolved in 600 μl of water, and neutralized with 180 μl of KOH (5 M) to pH 7, 76 mg of dithiothreitol dissolved in 500 μl of water, and 6.25 ml of $MgCl_2$ (1 M). To condition the native gel (3 to 8% Criterion XT tris-acetate protein gel, Bio-Rad), the gel was run in running buffer for 1 hour at 50 V at RT. After loading of the native samples, the gel was run at 30 mA for 5 to 6 hours in a cold room. The gel was then transferred onto a polyvinylidene difluoride membrane overnight at 20 V at 4°C before blocking and immunostaining against proteasome subunits. For the in-gel proteasome activity assay, 20 ml of peptide assay solution was prepared containing

1 ml of Tris-HCl, pH 7.5, with 100 μl of 1 M $MgCl_2$, 40 μl of 0.5 M ATP, and 100 μl of 10 mM Suc-LLVY-AMC (Abcam). The native gel was incubated with the assay solution for 1 hour at 37°C before washing with water and visualization using a Azure 280 gel imaging system (excitation, 365 nm).

For the in vitro enzymatic reaction between purified 19*S* and polyubiquitin chains, purified native cow 19*S* proteasome protein (ab218004) and purified human 26*S* proteasome protein (from HEK293; ab218006) were purchased from Abcam. The identities of the purified 19*S* and 26*S* particles were validated by native protein gel electrophoresis (fig. S1B). They were used at a final concentration of 12.5 nM. Recombinant human poly-ub wild-type chains (two to seven) Lys^{63} and Lys^{48} were purchased from R&D Systems (UCB330) and stored as a 1 mg/ml stock solution in water in 1:100 dilution. Then, 0.5 μl of recombinant proteasome particle stock solution was incubated with 0.5 μl of recombinant poly-ub chain stock solution and 19 μl of buffer containing 50 mM Tris, pH 7.4, 1 mM dithiothreitol, 5 mM $MgCl_2$, 2 mM ATP, and 250 mM sucrose for 2 hours at 37°C before SDS–polyacrylamide gel electrophoresis (SDS-PAGE). For the molecular weight ladder, we used the Odyssey One-Color Protein Molecular Weight Marker. From top to bottom, the bands of the ladder correspond to 250, 150, 100, 75, 50, 37, 25, 20, 15, and 10 kD.

Immunoprecipitation

Magnetic beads conjugated with K48 TUBE (UM407M, Tebu-bio), K63 (UM404M, Tebu-bio), and control magnetic beads (UM0500M, Tebu-bio) were used following an optimized protocol from the supplier. Briefly, four dishes of cortical neurons cultured in 10-cm petri dishes (4 million cells per dish) were scraped and pelleted in cold PBS before addition of 200 μl of lysis buffer containing 100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 100 μM PR619 (a ubiquitin protease inhibitor; Sigma), 5 mM o-phenanthroline (a DUB inhibitor; Sigma), and 5 mM *N*-ethylmaleimide (a DUB inhibitor; Merck). The pellets were homogenized in lysis buffer using a 1-ml syringe before centrifugation for 20 min at 4°C, 16k G-force. Then, 30 μl of clarified lysates were saved for Western blotting. An 80-μl slurry of the magnetic beads was equilibrated using a Tris-buffered saline with Tween-20 buffer as described by the supplier's protocol (20 mM Tris-HCl, pH 8, 150 mM NaCl, 0.1% Tween-20) and added into the remaining 170 μl of clarified lysate for 2 hours incubation at 4°C on a shaker. The beads were collected, washed, and eluted using 45-μl SDS-PAGE loading buffer. The supernatants were saved as a control. Immunoprecipitation was verified by Western blotting against Lys^{48} and Lys^{63} ubiquitin. The mixture of immunoprecipitated proteins and SDS-PAGE loading buffer was prepared for MS.

MS sample preparation

The immunoprecipitated proteins in SDS-PAGE loading buffer were prepared for MS-based proteomics analysis as previously reported (*56*). Briefly, samples were mixed with 2× lysis buffer (10% SDS, 100 mM Tris, pH 7.55, with H_3PO_4 supplemented with cOmplete protease inhibitor cocktail) in a 1:1 ratio and subsequently reduced using dithiothreitol in a final concentration of 20 mM for 10 min at RT. For cysteine alkylation, the samples were incubated with iodoacetamide in a final concentration of 50 mM for 30 min at RT in the dark. Then, the samples were acidified using phosphoric acid in a final concentration of ~1.2%. Binding/wash buffer (90% methanol, 50 mM Tris, pH 7.1 with H_3PO_4) was added in a 1:7 lysate-to-buffer ratio to the acidified proteins. The protein suspension was loaded onto the filter of the S-trap (size: "micro"; ProtiFi) by centrifugation for 20 s at 4000*g* in 150-μl steps. Trapped proteins were washed with 150 μl of binding/wash buffer four times. Trypsin (1 μg; Promega) was added in 60 μl of 40 mM ammonium bicarbonate buffer. Digestion was performed overnight (~18 hours) at RT in a humidified chamber. For peptide collection, the filter was washed in three consecutive steps by centrifugation at 4000*g* for 40 s starting with 40 μl of digestion buffer and two 40-μl washes with 0.2% formic acid in MS-grade water. The eluted peptides were further processed by C18-based purification ("stage tips") according to previously published protocols (*57*) and dried in vacuo.

LC-MS analysis

Dried peptide samples were reconstituted in 5% acetonitrile (ACN), 95% water, and 0.1% formic acid (FA). Peptides were separated by nano-high-performance liquid chromatography (nano-HPLC; U3000 RSLCnano, Dionex). The samples were loaded and washed with loading buffer (2% ACN, 0.05% trifluoroacetic acid in water for 6 min at 6 μl/min) on a PepMap100 loading column (C18, 20 mm length, 75 μm inner diamater, 3 μm particle size; Thermo Scientific). Subsequently, peptides were separated by a gradient of water (buffer A: 100% H_2O and 0.1% FA) and acetonitrile (buffer B: 80% ACN, 20% H_2O, and 0.1% FA) with a constant flow rate of 250 nl/min on an analytical column with an integrated emitter (C18, 500 mm length, 75 μm inner diameter, 1.7 μm particle size; CoAnn Technologies) and heated to 55°C. The gradient was from 4 to 48% in buffer B for 90 min. All solvents were liquid chromatography–MS (LC-MS) grade and were purchased from Riedel-de Häen/Honeywell. Eluting peptides were analyzed in data-dependent acquisition mode on a Fusion Lumos mass spectrometer (Thermo Fisher Scientific) coupled to the nano-HPLC by a Nanoflex source. MS1 survey scans were acquired over a scan range of 350 to 1400 mass-to-charge ratio (m/z) in the Orbitrap detector [120 k resolution (R), 2e5 automatic gain control, and 50 ms maximum injection time). Sequence information was acquired by a "top-speed" MS2 method with a fixed cycle time of 2 s for the survey and after MS/MS scans. MS2 scans were generated from the most abundant precursor ions with a minimum intensity of 5e3 and charge states from 2 to 5. Selected precursors were isolated in the quadrupole using a 1.4-Da window and fragmented using higher-energy C-trap dissociation (HCD) at 30% normalized collision energy. For MS2 scans, an AGC of 1e4 and a maximum injection time of 300 ms were used. Resulting fragment ions were detected in the ion trap using the rapid scan rate. Dynamic exclusion was set to 30 s with a mass tolerance of 10 parts per million (ppm). Each sample was measured in duplicate LC–tanden MS (LC-MS/MS) runs.

LC-MS analysis of gel bands

Gel bands were cut from native PAGE gels and prepared for LC-MS analysis using an in-gel tryptic digestion kit (89871, Thermo Scientific). The procedure was prepared according to the manufacturer's protocol, except all solvent volumes were increased by threefold for native PAGE preparation. After digestion, the supernatant was mixed with 5 μl of 0.1% FA in MS-grade water and transferred into an MS vial. For LC-MS analysis, peptides were separated by a nano-HPLC (U3000 RSLCnano, Dionex). The samples were loaded and washed with loading buffer (2% ACN, 0.05% trifluoroacetic acid in MS-grade water) for 6 min at 6 μl/min on a PepMap100 loading column (C18, 20 mm length, 3 μm particle size; Thermo Scientific) and subsequently separated on an analytical column (C18, 50 cm length, 1.5 μm particle size; PepSep; heated to 55°C) by a linear gradient of solvent A (0.1% FA in MS-grade water) and solvent B (80% ACN, 0.1% FA in MS-grade water). The gradient ramped from 4 to 48% solvent B for 90 min at a flow rate of 300 nl/min. The eluting peptides were ionized using a Nanospray Flex ion source (Thermo Scientific) and then analyzed in a Qexactive Plus mass spectrometer (Thermo Scientific). Precursor ion spectra were acquired over the mass range of 350 to 1400 m/z (70 k mass resolution, 3e6 AGC target, 60 ms maximum injection time), and top-10 precursor ions were selected for fragmentation (HCD; 30% normalized collision energy) and analysis in MS2 mode (17.5 k mass resolution, 1.7 Da isolation window, 2.4e4 AGC target, 50 ms maximum injection time).

LC-MS data processing

For protein identification and quantification, MS raw data were processed using MaxQuant

(version 1.6.2.3; RRID: SCR_015753) with customized Andromeda parameters (58). For all searches, spectra were matched to a *Rattus norvegicus* sequence database (reviewed and unreviewed; downloaded from uniprot.org; RRID:SCR_004426) and to contaminant and decoy databases. Tryptic peptides with 0 to 2 missed cleavages were considered. Precursor mass tolerance was set to 4.5 ppm and fragment ion tolerance to 0.5 Da. Carbamidomethylation of cysteine residues was selected as a fixed modification, and protein N-terminal acetylation and methionine oxidation were selected as variable modifications. A false discovery rate (FDR) of 1% was applied at the peptide-spectrum-match (PSM) and protein levels. Proteins identified by at least one unique peptide were retained for downstream analysis. Label-free quantification (LFQ) was performed by pairwise ratio determination in at least three consecutive full scans; LFQ normalization was deselected. The match-between-runs (MBR) option was enabled. Downstream analysis of each pulled-down proteome was performed using Perseus (version 1.6.2.3; RRID: SCR_015753). First, decoy and contaminant proteins were discarded from the quantified protein groups (proteinGroups.txt). Proteins were then filtered on the basis of their uniqueness to the IP or fold enrichment (greater than threefold) compared with the "empty-bead" control. Gene Ontology (GO) term overrepresentation was analyzed comparing the gene names of the respective pulled-down proteins with gene names of the proteins detected in input samples (i.e., total neuronal cortex lysates) using a Fisher's exact test with Benjamini-Hochberg correction (FDR <1%). For gel-band analyses, protein quantification was performed selecting the intensity-based absolute quantification (iBAQ) option in MaxQuant.

Electrophysiology

Cultured cortical neurons (18 to 21 days in vitro) treated with dimethyl sulfoxide (DMSO) or b-AP15 (5 μM, 1 hour) were used for whole-cell recordings of miniature synaptic transmission in HEPES-buffered artificial cerebrospinal fluid (ACSF), as previously reported (59). Data were analyzed offline using Clampfit. After a blinded manual baseline correction, mEPSC events were screened with an amplitude threshold of ≥ 4 pA and an exponential decay. Slice electrophysiological experiments were performed in acute hippocampal brain slices prepared from 4- to 6-week-old male Sprague Dawley rats. Rats were anesthetized with isoflurane and decapitated with a guillotine after slowed breathing and a lack of movement and reflexes were assessed. Their brains were quickly removed from the skull (<1.5 min), put in ice-cold, slushed sucrose-based saline solution oxygenated with 95% O_2/5% CO_2, and cut into halves along the longitudinal fissure. Then, the neocortical side of the hemispheres were cut with an angle of 30°, glued on an ice-cold specimen holder, and sliced with a vibratome (VT 1200 S, Leica) to 400-μm-thick transverse slices in ice-cold, oxygenated sucrose-based saline solution. Before slice preparation, blades were adjusted parallel to the cutting surface using vibrocheck and set to a cutting speed of 0.08 mm/s with an amplitude of 1 mm. For experiments, slices from the middle part of the hippocampus were stored in a submerged slice holder in a beaker containing oxygenated ACSF19 at RT until the last section was collected and recovered for 1 hour submerged in the slice holder. The sucrose-based saline solution contained the following (in mM): 87 NaCl, 25 $NaHCO_3$, 2.3 KCl, 1.25 $NaH_2PO_4 \cdot H_2O$, 10 D-glucose, 75 sucrose, 0.5 $CaCl_2$, and 7 $MgCl_2$. ACSF contained the following (in mM): 119 NaCl, 26.2 $NaHCO_3$, 2.5 KCl, 1 $NaH_2PO_4 \cdot H_2O$, 11 D-glucose, 2.5 $CaCl_2$, and 1.3 $MgSO_4$. For measurements of fEPSPs, slices were transferred to an interface recording chamber on lens cleaning paper and continuously perfused with oxygenated ACSF at 31 ± 2°C at a speed of 2 ml/min for at least 1 hour before recordings. The Schaffer collateral pathway was stimulated every 30 s by the placement of a tungsten bipolar stimulation electrode positioned in stratum radiatum of area CA1 on one side of the glass recording electrode (3 to 4 MΩ) filled with oxygenated ACSF. Stimulation electrodes were lowered 200 μm deep into the slice, and the recording electrode was advanced to detect the largest possible response. Before baseline recordings, fEPSPs were measured for 20 to 60 min until they stabilized. After recording a 15-min baseline, b-AP15 (10 μM) or DMSO as a control was added to the perfusion medium.

Constructs and transfection

HA-wild-type ubiquitin and HA-K63R ubiquitin constructs were custom-made by IDT as gBlocks gene fragments and cloned in-house by Gibson assembly. Uch37/Uchl5 shRNAs, Usp14 shRNAS, and scrambled-sequence controls were purchased from OriGene (TL70097A-D, TL701657A-D, and TR30021, respectively). Lipofectamine 2000 (Thermo Fisher) was used for transfection following a modified protocol from the supplier. In brief, for each transfection, 2 μl of lipofectamine 2000 was mixed with 75 μl pof Neurobasal medium (supplemented with Glutamax) and incubated for 5 min at RT. In a separate tube, 1 μg of DNA was mixed with 75 μl of Neurobasal medium (supplemented with Glutamax). The lipofectamine-containing mixture was added into the DNA-containing mixture before mixing and 20 min incubation at RT. The original medium of cultured neurons was removed and saved before 150 μl of transfection mixture was added on the cells grown on Mattek dishes and incubated for 1 hour at 37°C, 5% CO_2. After incubation, the transfection mixture was aspirated from the cell dishes and the original medium was placed back in the dishes. Expression of constructs was checked after 2 days.

Reagents

Unless noted otherwise, all substances were molecular biology or cell culture grade and purchased from Sigma-Aldrich or Roth. MG132 (Merck) was stored as a 50 mM stock solution in DMSO and used at a 1:1000 dilution. Bortezomib was stored as a 2 mM stock solution in DMSO and used at a 1:2000 dilution. b-AP15 was purchased from Selleckchem (S4920), stored as a 10 mM stock solution in DMSO) and used at 5 μM. Capzimin was stored as a 5 mM stock colution in DMSO and used at a 1:1000 dilution.

Statistics

Comparisons between two groups were conducted using two-sample *t* tests. ANOVA with Bonferroni *post hoc* analysis was used for comparisons of more than two groups. $P < 0.05$ was considered significant for statistical analyses. For confocal microscopy that involved comparisons between groups, unless otherwise noted, at least 10 cell replicates from three preparations were measured for each group. For single-molecule localization microscopy, at least six cell replicates from three preparations were measured. For biochemistry experiments, at least three biological replicates were used for each group. For electrophysiology experiments, 16 recordings from at least four biological replicates were used for mEPSC measurements, and five biological replicates were used for fEPSP measurements. Box plots include the median (50th percentile) indicated by the middle line in the box, the interquartile range (25th to 75th percentile) indicated by the top and bottom outlines of the box, and the maximum and minimum without outliers (outliers defined as >75th percentile plus 1.5-fold of the interquartile range or <25th percentile minus 1.5-fold of the interquartile range). Violin plots include the mean values indicated by the middle line, the interquartile range (25th to 75th percentile) indicated by the black column, and the maximum and minimum without outliers indicated by the whiskers.

REFERENCES AND NOTES

1. C. E. Holt, K. C. Martin, E. M. Schuman, Local translation in neurons: Visualization and function. *Nat. Struct. Mol. Biol.* **26**, 557–566 (2019). doi: 10.1038/s41594-019-0263-5; pmid: 31270476
2. S. L. Giandomenico, B. Alvarez-Castelao, E. M. Schuman, Proteostatic regulation in neuronal compartments. *Trends Neurosci.* **45**, 41–52 (2022). doi: 10.1016/j.tins.2021.08.002; pmid: 34489114
3. C. Sun, E. M. Schuman, Logistics of neuronal protein turnover: Numbers and mechanisms. *Mol. Cell. Neurosci.* **123**, 103793 (2022). doi: 10.1016/j.mcn.2022.103793; pmid: 36396040
4. A.-S. Hafner, P. G. Donlin-Asp, B. Leitch, E. Herzog, E. M. Schuman, Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments. *Science* **364**, eaau3644 (2019). doi: 10.1126/science.aau3644; pmid: 31097639

5. C. Sun *et al*., The prevalence and specificity of local protein synthesis during neuronal synaptic plasticity. *Sci. Adv.* **7**, eabj0790 (2021). doi: 10.1126/sciadv.abj0790; pmid: 34533986
6. B. Bingol, E. M. Schuman, Activity-dependent dynamics and sequestration of proteasomes in dendritic spines. *Nature* **441**, 1144–1148 (2006). doi: 10.1038/nature04769; pmid: 16810255
7. A. Biever *et al*., Monosomes actively translate synaptic mRNAs in neuronal processes. *Science* **367**, eaay4991 (2020). doi: 10.1126/science.aay4991; pmid: 32001627
8. S. M. Asano *et al*., Expansion microscopy: Protocols for imaging proteins and RNA in cells and tissues. *Curr. Protoc. Cell Biol.* **80**, e56 (2018). doi: 10.1002/cpcb.56; pmid: 30070431
9. C. M. Fusco *et al*., Neuronal ribosomes exhibit dynamic and context-dependent exchange of ribosomal proteins. *Nat. Commun.* **12**, 6127 (2021). doi: 10.1038/s41467-021-26365-x; pmid: 34675203
10. T. Shigeoka *et al*., On-Site Ribosome Remodeling by Locally Synthesized Ribosomal Proteins in Axons. *Cell Rep.* **29**, 3605–3619.e10 (2019). doi: 10.1016/j.celrep.2019.11.025; pmid: 31825839
11. H. M. Goodman, A. Rich, Mechanism of polyribosome action during protein synthesis. *Nature* **199**, 318–322 (1963). doi: 10.1038/199318a0; pmid: 14058566
12. Z. Sha, J. Zhao, A. L. Goldberg, "Measuring the overall rate of protein breakdown in cells and the contributions of the ubiquitin-proteasome and autophagy-lysosomal pathways," in *The Ubiquitin Proteasome System*, T. Mayor, G. Kleiger, Eds. (Springer, 2018), vol. 1844 of *Methods in Molecular Biology*, pp. 261–276; http://link.springer.com/10.1007/978-1-4939-8706-1_17.
13. G. N. Patrick, B. Bingol, H. A. Weld, E. M. Schuman, Ubiquitin-mediated proteasome activity is required for agonist-induced endocytosis of GluRs. *Curr. Biol.* **13**, 2073–2081 (2003). doi: 10.1016/j.cub.2003.10.028; pmid: 14653997
14. R. Fonseca, R. M. Vabulas, F. U. Hartl, T. Bonhoeffer, U. V. Nägerl, A balance of protein synthesis and proteasome-dependent degradation determines the maintenance of LTP. *Neuron* **52**, 239–245 (2006). doi: 10.1016/j.neuron.2006.08.015; pmid: 17046687
15. H.-C. Tai, E. M. Schuman, Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. *Nat. Rev. Neurosci.* **9**, 826–838 (2008). doi: 10.1038/nrn2499; pmid: 18931696
16. G. A. Collins, A. L. Goldberg, The logic of the 26*S* proteasome. *Cell* **169**, 792–806 (2017). doi: 10.1016/j.cell.2017.04.023; pmid: 28525752
17. Y. T. Kwon, A. Ciechanover, The ubiquitin code in the ubiquitin-proteasome system and autophagy. *Trends Biochem. Sci.* **42**, 873–886 (2017). doi: 10.1016/j.tibs.2017.09.002; pmid: 28947091
18. E. B. Dammer *et al*., Polyubiquitin linkage profiles in three models of proteolytic stress suggest the etiology of Alzheimer disease. *J. Biol. Chem.* **286**, 10457–10465 (2011). doi: 10.1074/jbc.M110.149633; pmid: 21278249
19. J. A. Nathan, H. T. Kim, L. Ting, S. P. Gygi, A. L. Goldberg, Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes? *EMBO J.* **32**, 552–565 (2013). doi: 10.1038/emboj.2012.354; pmid: 23314748
20. F. Kriegenburg *et al*., Mammalian 26*S* proteasomes remain intact during protein degradation. *Cell* **135**, 355–365 (2008). doi: 10.1016/j.cell.2008.08.032; pmid: 18957208
21. K. B. Hendil, R. Hartmann-Petersen, K. Tanaka, 26 S proteasomes function as stable entities. *J. Mol. Biol.* **315**, 627–636 (2002). doi: 10.1006/jmbi.2001.5285; pmid: 11812135
22. T. Geiger, A. Wehner, C. Schaab, J. Cox, M. Mann, Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **11**, 014050 (2012). doi: 10.1074/mcp.M111.014050; pmid: 22278370
23. S. Asano *et al*., A molecular census of 26*S* proteasomes in intact neurons. *Science* **347**, 439–442 (2015). doi: 10.1126/science.1261197; pmid: 25613890
24. A. R. Dörrbaum, L. Kochen, J. D. Langer, E. M. Schuman, Local and global influences on protein turnover in neurons and glia. *eLife* **7**, e34202 (2018). doi: 10.7554/eLife.34202; pmid: 29914620
25. I. Sahu *et al*., The 20*S* as a stand-alone proteasome in cells can degrade the ubiquitin tag. *Nat. Commun.* **12**, 6173 (2021). doi: 10.1038/s41467-021-26427-0; pmid: 34702852
26. K. V. Ramachandran, S. S. Margolis, A mammalian nervous-system-specific plasma membrane proteasome complex that modulates neuronal function. *Nat. Struct. Mol. Biol.* **24**, 419–430 (2017). doi: 10.1038/nsmb.3389; pmid: 28287632
27. J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, R. Jungmann, Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017). doi: 10.1038/nprot.2017.024; pmid: 28518172
28. H.-C. Tai, H. Besche, A. L. Goldberg, E. M. Schuman, Characterization of the brain 26*S* proteasome and its interacting proteins. *Front. Mol. Neurosci.* **3**, 12 (2010). doi: 10.3389/fnmol.2010.00012; pmid: 20717473
29. S. A. H. de Poot, G. Tian, D. Finley, Meddling with fate: The proteasomal deubiquitinating enzymes. *J. Mol. Biol.* **429**, 3525–3545 (2017). doi: 10.1016/j.jmb.2017.09.015; pmid: 28988953
30. P. D'Arcy *et al*., Inhibition of proteasome deubiquitinating activity as a new cancer therapy. *Nat. Med.* **17**, 1636–1640 (2011). doi: 10.1038/nm.2536; pmid: 22057347
31. M. van Oostrum *et al*., The proteomic landscape of synaptic diversity across brain regions and cell types. bioRxiv 525780 [Preprint] (2023); https://10.1101/2023.01.27.525780.
32. G. K. Tofaris *et al*., Ubiquitin ligase Nedd4 promotes α-synuclein degradation by the endosomal-lysosomal pathway. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17004–17009 (2011). doi: 10.1073/pnas.1109356108; pmid: 21953697
33. S. L. Scudder *et al*., Synaptic strength is bidirectionally controlled by opposing activity-dependent regulation of Nedd4-1 and USP8. *J. Neurosci.* **34**, 16637–16649 (2014). doi: 10.1523/JNEUROSCI.2452-14.2014; pmid: 25505317
34. A. C. Giles, B. Grill, Roles of the HUWE1 ubiquitin ligase in nervous system development, function and disease. *Neural Dev.* **15**, 6 (2020). doi: 10.1186/s13064-020-00143-9; pmid: 32336296
35. J. Li *et al*., Capzimin is a potent and specific inhibitor of proteasome isopeptidase Rpn11. *Nat. Chem. Biol.* **13**, 486–493 (2017). doi: 10.1038/nchembio.2326; pmid: 28244987
36. M. D. Ehlers, Reinsertion or degradation of AMPA receptors determined by activity-dependent endocytic sorting. *Neuron* **28**, 511–525 (2000). doi: 10.1016/S0896-6273(00)00129-X; pmid: 11144360
37. J. Widagdo *et al*., Activity-dependent ubiquitination of GluA1 and GluA2 regulates AMPA receptor intracellular sorting and degradation. *Cell Rep.* **10**, 783–795 (2015). doi: 10.1016/j.celrep.2015.01.015; pmid: 25660027
38. E. M. Petrini *et al*., Endocytic trafficking and recycling maintain a pool of mobile surface AMPA receptors required for synaptic potentiation. *Neuron* **63**, 92–105 (2009). doi: 10.1016/j.neuron.2009.05.025; pmid: 19607795
39. K. Czöndör *et al*., Unified quantitative model of AMPA receptor trafficking at synapses. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3522–3527 (2012). doi: 10.1073/pnas.1109818109; pmid: 22331885
40. L. D. Cohen, N. E. Ziv, Neuronal and synaptic protein lifetimes. *Curr. Opin. Neurobiol.* **57**, 9–16 (2019). doi: 10.1016/j.conb.2018.12.007; pmid: 30677713
41. B. Alvarez-Castelao, E. M. Schuman, The regulation of synaptic protein turnover. *J. Biol. Chem.* **290**, 28623–28630 (2015). doi: 10.1074/jbc.R115.657130; pmid: 26453306
42. T. Grune *et al*., HSP70 mediates dissociation and reassociation of the 26*S* proteasome during adaptation to oxidative stress. *Free Radic. Biol. Med.* **51**, 1355–1364 (2011). doi: 10.1016/j.freeradbiomed.2011.06.015; pmid: 21767633
43. J. D. Perez *et al*., Subcellular sequencing of single neurons reveals the dendritic transcriptome of GABAergic interneurons. *eLife* **10**, e63092 (2021). doi: 10.7554/eLife.63092; pmid: 33404500
44. B. Fricke, S. Heink, J. Steffen, P.-M. Kloetzel, E. Krüger, The proteasome maturation protein POMP facilitates major steps of 20*S* proteasome formation at the endoplasmic reticulum. *EMBO Rep.* **8**, 1170–1175 (2007). doi: 10.1038/sj.embor.7401091; pmid: 17948026
45. E. Zavodszky, S.-Y. Peak-Chew, S. Juszkiewicz, A. J. Narvaez, R. S. Hegde, Identification of a quality-control factor that monitors failures during proteasome assembly. *Science* **373**, 998–1004 (2021). doi: 10.1126/science.abc6500; pmid: 34446601
46. K. Liu *et al*., PI31 is an adaptor protein for proteasome transport in axons and required for synaptic development. *Dev. Cell* **50**, 509–524.e10 (2019). doi: 10.1016/j.devcel.2019.06.009; pmid: 31327739
47. A. J. Giovannone *et al*., The H_{abc} domain of syntaxin 3 is a ubiquitin binding domain. *Sci. Rep.* **10**, 21350 (2020). doi: 10.1038/s41598-020-78412-0; pmid: 33288783
48. G. V. Rinetti, F. E. Schweizer, Ubiquitination acutely regulates presynaptic neurotransmitter release in mammalian neurons. *J. Neurosci.* **30**, 3157–3166 (2010). doi: 10.1523/JNEUROSCI.3712-09.2010; pmid: 20203175
49. J. H. Vaden *et al*., Ubiquitin-specific protease 14 regulates c-Jun N-terminal kinase signaling at the neuromuscular junction. *Mol. Neurodegener.* **10**, 3 (2015). doi: 10.1186/1750-1326-10-3; pmid: 25575639
50. A. Al-Shami *et al*., Regulators of the proteasome pathway, Uch37 and Rpn13, play distinct roles in mouse development. *PLOS ONE* **5**, e13654 (2010). doi: 10.1371/journal.pone.0013654; pmid: 21048919
51. J. A. Harrigan, X. Jacq, N. M. Martin, S. P. Jackson, Deubiquitylating enzymes and drug discovery: Emerging opportunities. *Nat. Rev. Drug Discov.* **17**, 57–78 (2018). doi: 10.1038/nrd.2017.152; pmid: 28959952
52. M. Békés, D. R. Langley, C. M. Crews, PROTAC targeted protein degraders: The past is prologue. *Nat. Rev. Drug Discov.* **21**, 181–200 (2022). doi: 10.1038/s41573-021-00371-6; pmid: 35042991
53. J. R. Kowalski, P. Juo, The role of deubiquitinating enzymes in synaptic function and nervous system diseases. *Neural Plast.* **2012**, 892749 (2012). doi: 10.1155/2012/892749; pmid: 23316392
54. A. J. George, Y. C. Hoffiz, A. J. Charles, Y. Zhu, A. M. Mabb, A comprehensive atlas of E3 ubiquitin ligase mutations in neurological disorders. *Front. Genet.* **9**, 29 (2018). doi: 10.3389/fgene.2018.00029; pmid: 29491882
55. L. Groc *et al*., Differential activity-dependent regulation of the lateral mobilities of AMPA and NMDA receptors. *Nat. Neurosci.* **7**, 695–696 (2004). doi: 10.1038/nn1270; pmid: 15208630
56. K. Desch, J. D. Langer, E. M. Schuman, Dynamic bi-directional phosphorylation events associated with the reciprocal regulation of synapses during homeostatic up- and down-scaling. *Cell Rep.* **36**, 109583 (2021). doi: 10.1016/j.celrep.2021.109583; pmid: 34433048
57. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007). doi: 10.1038/nprot.2007.261; pmid: 17703201
58. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008). doi: 10.1038/nbt.1511; pmid: 19029910
59. A. R. Dörrbaum, B. Alvarez-Castelao, B. Nassim-Assir, J. D. Langer, E. M. Schuman, Proteome dynamics during homeostatic scaling in cultured neurons. *eLife* **9**, e52939 (2020). doi: 10.7554/eLife.52939; pmid: 32238265
60. Y. Perez-Riverol *et al*., The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022). doi: 10.1093/nar/gkab1038; pmid: 34723319

ACKNOWLEDGMENTS

We thank A. Schwarz for help with transmission electron microscopy; C. Fusco for help with compartmentalized chambers; A. Nold for help with Python codes; D. Vogel for help with DNA-antibody conjugation; T. Blok and M. van Oostrum for help with brain slices; M. Heilemann for helpful discussions; I. Bartnik, N. Fuerst, A. Staab, and C. Thum for cell culture support; and the staff of the imaging facility, proteomics facility, and scientific computing unit at the Max Planck Institute for Brain Research. **Funding:** This work was supported by an EMBO Long-Term Fellowship (EMBO ALTF 860-2018 to C.S.); an HFSP Cross-Disciplinary Fellowship (LT000737/2019-C to C.S.); a Joachim Herz Stiftung Add-on Fellowship for Interdisciplinary Life Science (850027 to C.S.); an FEBS Long-Term Fellowship (S.L.G.); Hessen Horizon, Marie Skłodowska-Curie-Stipendium Hessen (S.L.G.); the Max Planck Society (C.S., K.D., B.N.-A., S.L.G., P.N., J.D.L., E.M.S.); an Advanced Investigator Award from the European Research Council (743216 to E.M.S.); the German Research Foundation (DFG CRC 1080: "Molecular and Cellular Mechanisms of Neural Homeostasis" and DFG CRC 902: "Molecular Principles of RNA-based Regulation" to E.M.S.). **Author contributions:** Conceptualization: C.S., E.M.S.; Funding acquisition: J.L., E.M.S.; Investigation: C.S., K.D., B.N.-A., S.L.G., P.N., J.D.L., E.M.S.; Methodology: C.S., K.D., B.N.-A., S.L.G., J.D.L., E.M.S.; Project administration: E.M.S.; Supervision: E.M.S.; Visualization: C.S., E.M.S.; Writing – original draft: C.S., E.M.S.; Writing – review and editing: C.S., K.D., B.N.-A., S.L.G., P.N., J.D.L., E.M.S. **Competing interests:** The authors declare no competing interests. **Data and**

materials availability: All proteomics data associated with this manuscript have been uploaded to the PRIDE online repository (*60*) and can be accessed through dataset identifiers PXD037893 and PXD041161. All other data are available in the main text or the supplementary materials. Minimum Information for Biological and Biomedical Investigations (MIBBI) guidelines have been followed. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2018
Figs. S1 to S10
Data S1 and S2
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 5 October 2022; resubmitted 17 February 2023
Accepted 4 April 2023
10.1126/science.adf2018

RESEARCH ARTICLE

OCEAN NUTRIENTS

Natural iron fertilization by shallow hydrothermal sources fuels diazotroph blooms in the ocean

Sophie Bonnet[1]*, Cécile Guieu[2]*, Vincent Taillandier[2], Cédric Boulart[3], Pascale Bouruet-Aubertot[4], Frédéric Gazeau[2], Carla Scalabrin[5], Matthieu Bressac[2], Angela N. Knapp[6], Yannis Cuypers[4], David González-Santana[7,8], Heather J. Forrer[6], Jean-Michel Grisoni[9], Olivier Grosso[1], Jérémie Habasque[7], Mercedes Jardin-Camps[1], Nathalie Leblond[9], Frédéric A. C. Le Moigne[1,7], Anne Lebourges-Dhaussy[7], Caroline Lory[1], Sandra Nunige[1], Elvira Pulido-Villena[1], Andrea L. Rizzo[10,11], Géraldine Sarthou[7], Chloé Tilliette[2]

Iron is an essential nutrient that regulates productivity in ~30% of the ocean. Compared with deep (>2000 meter) hydrothermal activity at mid-ocean ridges that provide iron to the ocean's interior, shallow (<500 meter) hydrothermal fluids are likely to influence the surface's ecosystem. However, their effect is unknown. In this work, we show that fluids emitted along the Tonga volcanic arc (South Pacific) have a substantial impact on iron concentrations in the photic layer through vertical diffusion. This enrichment stimulates biological activity, resulting in an extensive patch of chlorophyll (360,000 square kilometers). Diazotroph activity is two to eight times higher and carbon export fluxes are two to three times higher in iron-enriched waters than in adjacent unfertilized waters. Such findings reveal a previously undescribed mechanism of natural iron fertilization in the ocean that fuels regional hotspot sinks for atmospheric CO_2.

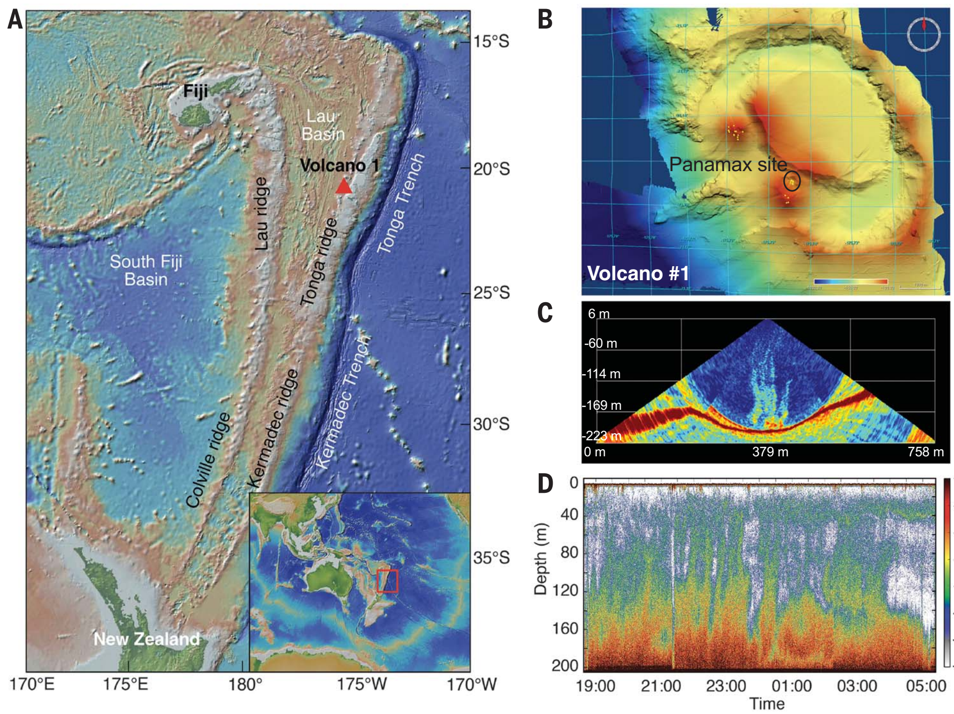
Planktonic diazotrophs are microscopic organisms ubiquitous in the ocean that play a crucial role: They supply available nitrogen (N), an essential but scarce nutrient in most of the oceans, to the surface ocean biosphere (*1*, *2*). Diazotrophs do so by converting atmospheric N_2 (endlessly available but metabolically useless) to ammonia (readily bioavailable), a reaction termed biological N_2 fixation. Diazotrophs thus alleviate N limitation in 60% of the oceans, especially in low latitudes. This promotes CO_2 fixation by phytoplankton into organic carbon (primary productivity) that, in turn, sustains the food web and organic carbon export and sequestration to the deep ocean (*3*–*6*). However, diazotrophs face a major challenge: Besides phosphorus requirements, the iron (Fe)–rich nitrogenase enzyme that catalyzes N_2 fixation imposes a high Fe demand on diazotroph growth (*7*), but Fe bioavailability in the ocean often limits the growth of these organisms (*6*, *8*). The western tropical South Pacific (WTSP) Ocean is a recognized hotspot of N_2 fixation activity, with an estimated contribution of ~21% to the global fixed N input (*9*). Fe supply through atmospheric deposition is known to control large-scale diazotroph biogeography (*10*), but such aeolian inputs are extremely low in this remote region (*11*), which suggests that alternative Fe fertilization processes underlie the ecological success of diazotrophs. Identifying these processes is of the utmost importance because diazotrophs have recently been identified as key drivers of future marine net primary productivity in response to climate change (*12*). In this work, we demonstrate that Fe-rich fluids emitted by shallow hydrothermal venting directly fertilize the overlying surface ecosystem, which induces intense diazotroph activity that supports enhanced carbon export fluxes, with a C sequestration efficiency higher than those from artificial mesoscale Fe-addition experiments.

The WTSP hosts the Tonga-Kermadec subduction zone, which stretches 2500 km from New Zealand to Tonga (Fig. 1A). It is the fastest-converging, most seismically active subduction zone and has the highest density of underwater volcanic centers on Earth (*13*). This system produces extensive plumes of ^{3}He in the bathypelagic ocean (1500 to 2000 m) that fingerprint deep hydrothermal sources originating in Lau Basin (*14*, *15*). Other authors (*16*, *17*) also identified shallower sources (<500 m) along the Tonga arc, which are associated with elevated dissolved Fe (DFe) and manganese (DMn) concentrations close to the seafloor. Guieu *et al.* (*11*) demonstrated that these shallow sources were able to bring DFe up to the photic layer (~100 m) at high concentrations (up to 66 nmol liter^{-1}). These Fe infusions are hypothesized to fuel the observed N_2 fixation hotspot associated with a patch of increased chlorophyll that persists 6 months per year in this region (Fig. 2, A and B) (*9*, *11*). Yet, there is, at present, no empirical evidence of the direct effect of such hydrothermal Fe fertilization on the overlying planktonic ecosystem, with the implication that a substantial part of N entering the subtropical Pacific, owing to hydrothermal Fe, is likely missing from N budgets. Such an Fe supply mechanism would challenge the prevailing paradigm that diazotroph productivity is mainly mediated by Fe from dust deposition (*10*) in N-limited regions.

To document the mechanistic link between Fe supply from submarine volcanism and the response of the surface plankton community, we combined acoustic, chemical, physical, and biological data acquired during the GEOTRACES GPpr14 TONGA expedition (*18*), a zonal transect between the Tonga volcanic arc and the South Pacific Gyre, which serves as a reference deep-sea site where the ocean surface is not affected by hydrothermal activity. The targeted submarine volcano [volcano 1 (*16*, *17*)] is a large stratovolcano (basal diameter 28 km) located in the central part of the Tonga arc (21°09.24′S; 175°44.64′W) (Fig. 1, A and B). During an acoustic survey above the volcano, we detected multiple acoustic plumes (Fig. 1B and table S1) rising from the sea floor up to ~20 m below the ocean surface (Fig. 1C). We focused our study on a site located near the caldera on the southwestern edge of the volcano (Fig. 1B) (hereafter referred to as the "Panamax site"), where the acoustic anomaly (*19*–*21*), which is also associated with intense gas-bubble emissions, was strong and continuous (Fig. 1D). Repeated conductivity, temperature, and depth (CTD) casts at this site revealed that acoustic plumes were also associated with strong anomalies in pH, turbidity, and redox potential (Eh) (Fig. 1E and table S2) from the seafloor (195 m depth) up to ~160 m. Methane concentrations that reached >100 nmol liter^{-1} (Fig. 1F) and the excess of ^{3}He and ^{4}He concentrations (fig. S1 and table S3) confirmed the hydrothermal origin of the plumes.

DFe and DMn were also enriched ~80-fold at this site (Fig. 1F) compared with similar depths in the WTSP (*11*, *22*). DFe and methane (CH_4) concentrations were positively correlated [coefficient of determination (R^2) = 0.89, $p < 0.05$]. DFe reached concentrations as high

[1]Aix Marseille University, Université de Toulon, CNRS, IRD, MIO Marseille, France. [2]Laboratoire d'Océanographie de Villefranche (LOV), Institut de la Mer de Villefranche, CNRS, Sorbonne Université, 06230 Villefranche-sur-Mer, France. [3]Adaptation et Diversité en Milieu Marin, UMR 7144 AD2M CNRS-Sorbonne Université, Station Biologique de Roscoff, 29680 Roscoff, France. [4]Laboratoire d'Océanographie et du Climat: Expérimentation et Approches Numériques (LOCEAN-IPSL), Sorbonne University, CNRS-IRD-MNHN, 75005 Paris, France. [5]Ifremer, Univ Brest, CNRS, UMR 6538 Geo-Ocean, F-29280 Plouzané, France. [6]Department of Earth, Ocean, and Atmospheric Sciences, Florida State University, Tallahassee, FL 32306, USA. [7]CNRS, Univ Brest, IRD, Ifremer, UMR 6539, LEMAR, Plouzané, France. [8]Instituto de Oceanografía y Cambio Global (IOCAG), Universidad de Las Palmas de Gran Canaria, 35017 Las Palmas, Spain. [9]Institut de la Mer de Villefranche, IMEV, Sorbonne Université, Villefranche-sur-Mer, France. [10]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Milano, Via Alfonso Corti 12, 20133 Milano, Italy. [11]Department of Earth and Environmental Sciences, University of Milano-Bicocca, Piazza della Scienza 4, 20126 Milan, Italy.
***Corresponding author. Email: sophie.bonnet@mio.osupytheas.fr (S.B.); cecile.guieu@imev-mer.fr (C.G.)**



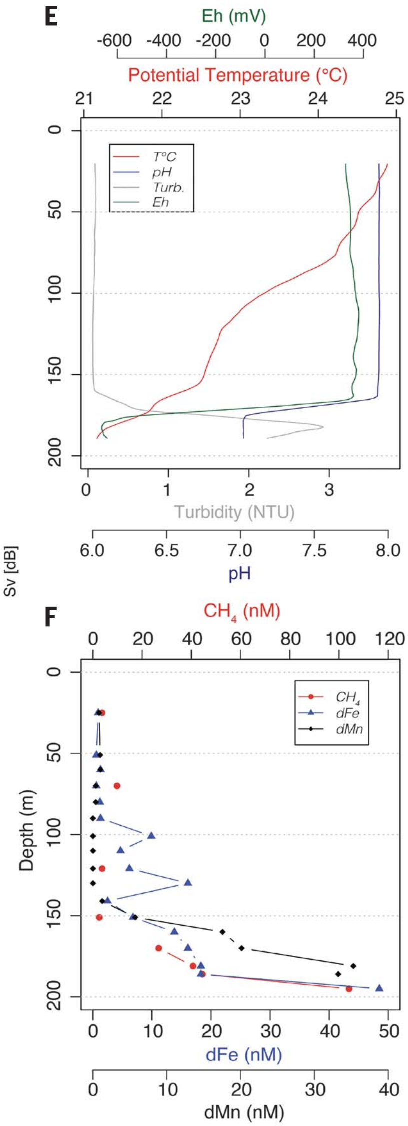


Fig. 1. Acoustic, optical, and chemical anomalies measured above volcano 1. (**A**) Location map showing the Tonga volcanic arc system and volcano 1 (red triangle, 21°09.24′S; 175°44.64′W). The inset shows a zoomed-out view of the region, with the area of the location map outlined with a red rectangle. (**B**) Bathymetry of volcano 1 performed during TONGA and position of the fluid and gas active sites detected by the multibeam echo sounder (EM710, 70 to 110 kHz) during the survey. The black circle represents the specific site studied here (Panamax site), from which the data shown in (C) to (F) were obtained. (**C**) Multibeam echo sounder image (EM710, 70 to 110 kHz) showing hydrothermal gas and fluid emissions from the seafloor rising up to ~10 m below the sea surface. (**D**) Time series (11 hours) of acoustic signal detected by the sounder EK60 (38 kHz) showing gas and fluid seafloor emissions at a fixed position [visualization threshold (Sv) of 75 dB]. (**E**) Vertical CTD profiles of temperature, pH, turbidity, and Eh in the main acoustic signal. NTU, nephelometric turbidity unit. (**F**) Vertical profiles of CH_4 (nM), DMn (nM), and DFe (nM) concentrations above volcano 1.

as 48.5 nmol liter^{-1} at 195 m (within the main acoustic signal), and although they decreased toward the photic layer (~0 to 100 m), DFe concentrations in that layer (~0.6 to 10 nmol liter^{-1}) (Fig. 1F) were one order of magnitude higher compared with those at stations not affected by hydrothermal activity (*23*). The turbulence profiles (fig. S2) revealed an order of magnitude higher vertical diffusivity above the volcano ($K_z = 3.7 \pm 1.9 \times 10^{-5}$ m^2 s^{-1} at ~50 m, which corresponds to the base of the surface mixed layer) compared with the distal open-sea reference site ($K_z = 5.2 \pm 9.6 \times 10^{-6}$ m^2 s^{-1}) (Table 1), in line with previous work above shallow (~200 m) seamounts (*24*). Combining the measured K_z with the DFe gradients (Table 1 and supplementary materials), the diffusive DFe vertical supply to the mixed layer above the volcano reached $1.1 \pm 1.7 \times 10^{-4}$ mmol Fe m^{-2} day^{-1}. This is orders of magnitude larger than the value at the reference site (Table 1), which suggests that Fe-rich fluids released close to the shallow volcano represent a major Fe source to surface waters. A phosphate supply of $5.4 \pm 2.4 \times 10^{-3}$ mmol m^{-2} day^{-1} accompanied this vertical DFe supply, whereas no nitrate supply could be quantified (table S5) because of the decoupling between the depth of the phosphacline (~50 m) and the nitracline (~100 m) (fig. S5), as already observed in the WTSP (*25*).

Along the west-to-east zonal transect, total chlorophyll a (Chla) and particulate organic N stocks peaked in the naturally Fe-fertilized waters at volcano 1 (Fig. 2). Both were also elevated up and downstream of the arc (Fig. 2, C and D), consistent with ocean color images (Fig. 2A). This biomass peak was associated with N_2 fixation rates enhanced two- to eightfold relative to surrounding waters on either side of the arc ($p < 0.05$, Mann-Whitney test) (Fig. 2E) and 90-fold higher *Trichodesmium* spp. abundances ($p < 0.05$, Mann-Whitney test) (Fig. 2F). This led to extremely high N_2 fixation rates and *Trichodesmium* spp. abundances in the Fe-fertilized waters (>2000 µmol N m^{-2} day^{-1}; ~6×10^7 *nifH* gene copies liter^{-1}, where *nifH* is the gene that encodes a subunit of the nitrogenase enzyme that catalyzes N_2 fixation), that is, values one to two orders of magnitude greater than those commonly found in other (sub)tropical ocean basins (*26*). This peak in diazotroph activity was marked by a phosphate drawdown (~50 nM) in the photic layer, although concentrations were not limiting for *Trichodesmium* spp. (*27*) because of intense microbial phosphorus cycling in this region (*28*). Diazotrophs were favored by the extremely low nitrate concentrations along the transect (fig. S6).

The particulate organic carbon (POC) and nitrogen (PON) export fluxes were measured using surface tethered sediment traps deployed for 4 days near volcano 1 and at the reference site. Consistent with model simulations in this region (*22*), POC export at 170

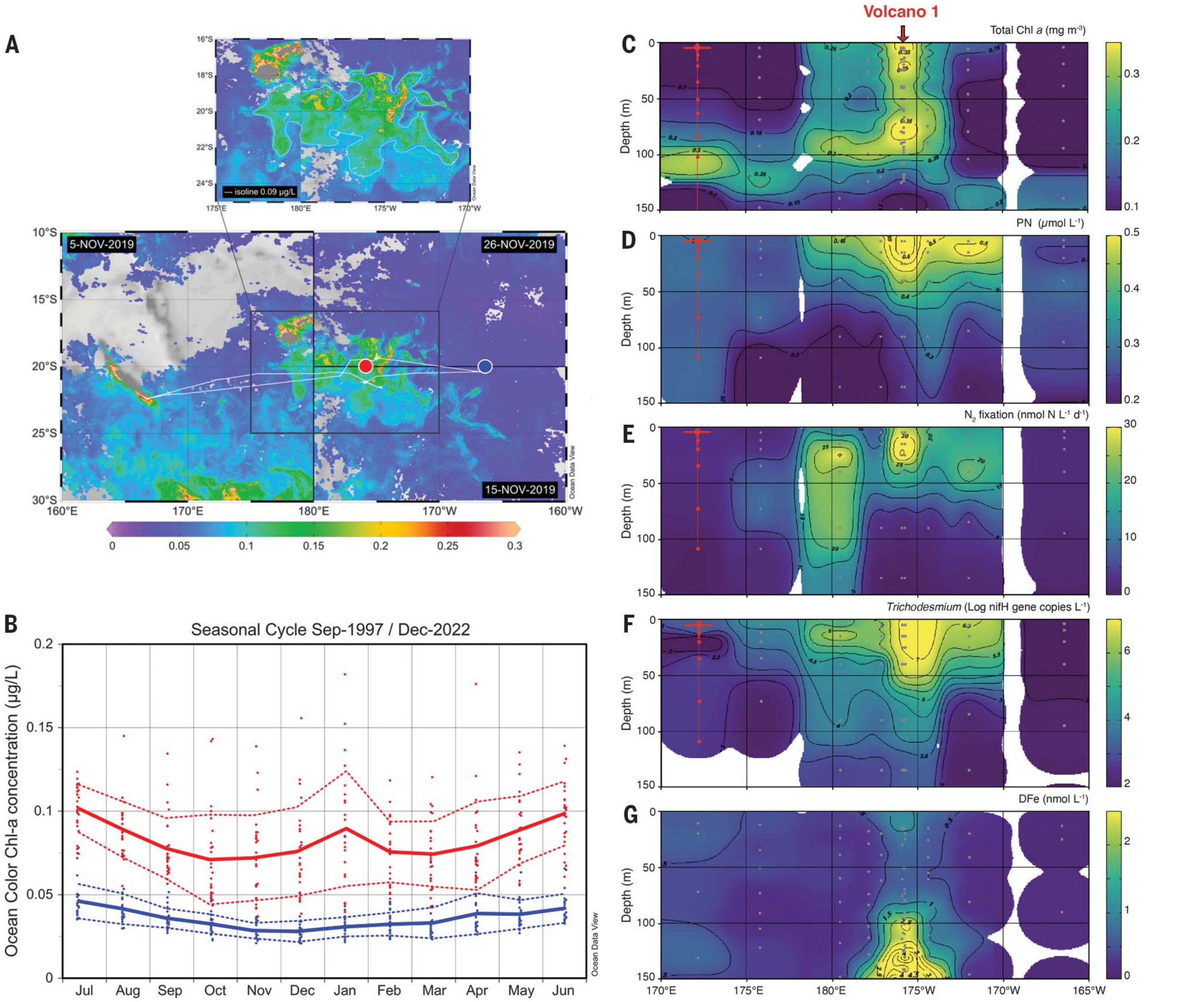


Fig. 2. Chlorophyll patch in the vicinity of Tonga and associated biogeochemical and biological parameters. (**A**) Surface [Chla] [GlobColour chlorophyll-a product based on multisensor satellite observations (SeaWiFS, MERIS, MODIS, VIIRS, and OLCI)] averaged over the time period corresponding to the TONGA cruise (1 November to 6 December 2019) at a resolution of 4 km and zoom-in of the bloom region (~360,000 km^2). The bloom is delineated by the isoline 0.9 μg liter^{-1} (which corresponds to twice the average background [Chla] outside the bloom). The red and blue circles represent the reference sites for the data shown in (B). (**B**) Monthly climatology of [Chla] from a 25-year time series [GlobColour chlorophyll-a product based on multisensor satellite observations (SeaWiFS, MERIS, MODIS, VIIRS, and OLCI)] at two sites along 20°S; the solid red line represents the multiyear average at monthly resolution in the region of the Tonga bloom represented with the red circle in (A), and the solid blue line corresponds to the location of the reference site in the South Pacific Gyre, represented as a blue circle in (A). The red and blue dashed lines represent the standard deviation. (**C** to **G**) Horizontal and vertical distributions of (C) total Chla concentrations (μg liter^{-1}), (D) particulate organic nitrogen (PN) concentrations (μmol liter^{-1}), (E) N_2 fixation rates (nmol N liter^{-1} day^{-1}), (F) *Trichodesmium* abundances (log *nifH* gene copies liter^{-1}), and (G) DFe concentrations (nmol liter^{-1}). The *y* axis shows depth (m), and the *x* axis shows longitude. Gray dots correspond to sampling depths at the various stations.

and 270 m was two to three times higher in the Fe-fertilized patch than at the reference site (Table 1), which resulted in an excess of POC export of 1.5 to 2.5 mmol C m^{-2} day^{-1} in the fertilized waters. Comparing measurements of subsurface water column nitrate plus nitrite isotope ratios (δ^{15}N) [1.2 to 2.2 per mil (‰)] with the δ^{15}N of sinking PON (−0.5 ± 3.5‰ at 170 m and −0.2 ± 1.9‰ at 270 m, respectively), the N isotope budget (N_2 fixation end member = −1‰) revealed that N_2 fixation supported 77 to 84 ± 159% at 170 m and 64 to 75 ± 86% at 270 m of the export production in the Fe-fertilized area, consistent with the massive export of diazotrophs observed in the traps during the expedition (*3*, *29*). Collectively, these results suggest that the hydrothermally driven Fe fertilization fuels planktonic diazotrophs, which results in the low δ^{15}N of sinking PON and high POC export fluxes compared with subtropical systems not affected by hydrothermal activity (table S6) (*30*).

To confirm the causal link between hydrothermal inputs and diazotroph activity, we conducted experiments in which hydrothermally enriched waters collected at the Panamax site of volcano 1 (hereafter referred to as "plume water") were supplied to surface biological communities by using 300-liter reactors free of trace metals (fig. S7A and methods). Increasing volumes of plume water (from 0 to 14.5% of the total reactor volume) were added to surface seawater from outside of the direct volcanic

influence (21°40.98′S; 174°42.54′W). The plume water was characterized by low pH (6.4) compared with ambient seawater (8.1), and DFe concentrations (15.8 nmol liter^{-1}) were ~15-fold higher than those in ambient surface waters (1.0 nmol liter^{-1}) (table S7). The increasing additions resulted in consistent increases in DFe concentration and decreasing pH in the experimental reactors (fig. S7). Plume water additions enhanced N_2 fixation rates by a factor of seven to eight, on average, over all sampling days compared with those measured in the unamended control ($p < 0.05$, Mann-Whitney test) (Fig. 3A) and reached levels in the same range as in situ rates measured above the volcano (Fig. 2E). Likewise, as observed in situ, *Trichodesmium* abundances increased by a factor of three- to fivefold ($p < 0.05$, Mann-Whitney test) in plume water–amended reactors (Fig. 3B). Both N_2 fixation rates and *Trichodesmium* abundances decreased at the end of the experiment, likely as a consequence of phosphate depletion in the closed reactors (no possible turbulent diffusion) (fig. S7), but generally remained higher in the amended reactors compared with those measured in the control. Because of the low inorganic N:P molar ratio (~9:1) in the plume water (indicative of a greater phosphate availability relative to nitrate), the mixing of the plume water with surface seawater likely prevented the phosphate limitation for up to days 6 to 8, whereas nitrate was depleted after 2 days. This decoupling between nitrate and phosphate also mirrors the in situ data described above.

The Fe supply from the Tonga arc thus drives, in large part, the upper ocean phenology of biological activity. We estimate that the region of elevated Chla extends ~800 km in longitude and ~450 km in latitude (Fig. 2A), forming a hotspot of biological activity of ~360,000 km^2 in the middle of the otherwise desert-like WTSP. The trajectories of surface velocity program drifters deployed above volcano 1 indicate that over a 6-month period, Fe-fertilized water masses can be dispersed regionally and support this extended Chla patch (fig. S8). The trajectories provide a bulk representation of the complex dynamical processes occurring at smaller scales that involve the South Equatorial current modulated by mesoscale activity (*11*, *31*) or lateral stirring by filaments (*32*). In addition, multiple active vent fields have been (recently) identified along the Tonga arc and the Lau Basin (*16*, *17*, *33*, *34*), either at shallow depths (<500 m) or deeper (500 to 1000 m). Although all active shallow vents have not yet been discovered, with a density estimation of one active volcano center per 12 km of arc, fertilization processes such as those evidenced at volcano 1 likely occur at many locations along the arc, which further explains the regional extent of the Chla patch observed by satellite (Fig. 2A). We cannot exclude that the few emerged Tonga islands could provide additional nutrients, likely contributing to the observed bloom. However, a study conducted over the entire tropical Pacific reveals that the mass effects of these islands are generally very localized (around the islands, with a Chla patch area of 9 to 13 km^2) and are of moderate amplitude (+9% of Chla increase relative to background waters) (*35*). By contrast, the bloom in the region of the Tonga volcanic arc is distinctive in that it is much larger (360,000 km^2) and of greater amplitude (~100% Chla increase) than around any other Pacific islands or archipelagos. This means that additional Fe sources of hydrothermal origin are necessary to sustain such a bloom. As an example, very weak Chla is observed around the Cook Islands, which are located at the same latitude as Tonga and receive the same amount of rainfall and aerosol Fe deposition annually but are not affected by hydrothermal activity (fig. S9), which confirms that nutrients of terrestrial origin are not sufficient to sustain blooms of large amplitude. Lastly, some extremely rare events such as the massive eruption of the Hunga Tonga Hunga Ha'apai volcano in January 2022 could also cause short-term localized blooms, such as the one observed by satellite after the eruption (*36*), that had no visible effect on the interannual trend of Chla. Other authors suggest that the interpretation

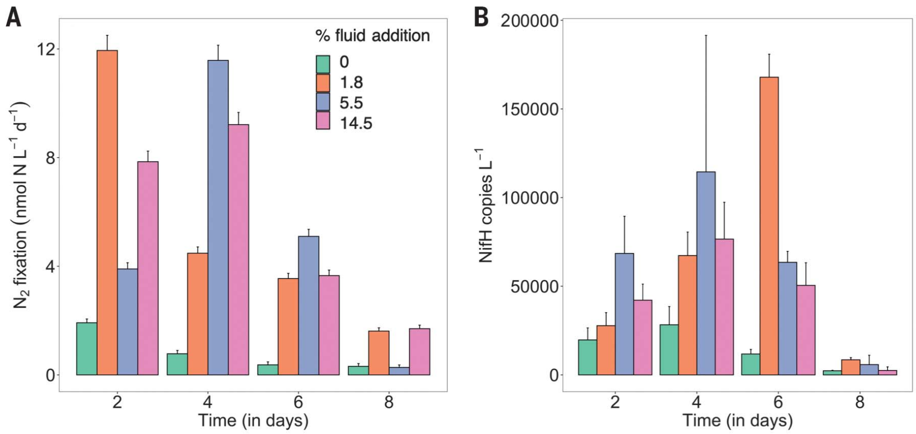


Fig. 3. Experimental evidence of the impact of hydrothermally enriched water additions on diazotroph communities. (**A** and **B**) Temporal evolution of (A) N_2 fixation rates (nmol N liter^{-1} day^{-1}) and (B) *Trichodesmium* abundances (*nifH* gene copies liter^{-1}) over the 196 hours of the experiment in the control treatment (no fluid addition) and in the reactors amended with 1.8, 5.5, and 14.5% of plume water collected at the Panamax site of volcano 1 (~200 m depth). Error bars correspond to standard deviations on triplicate analyses.

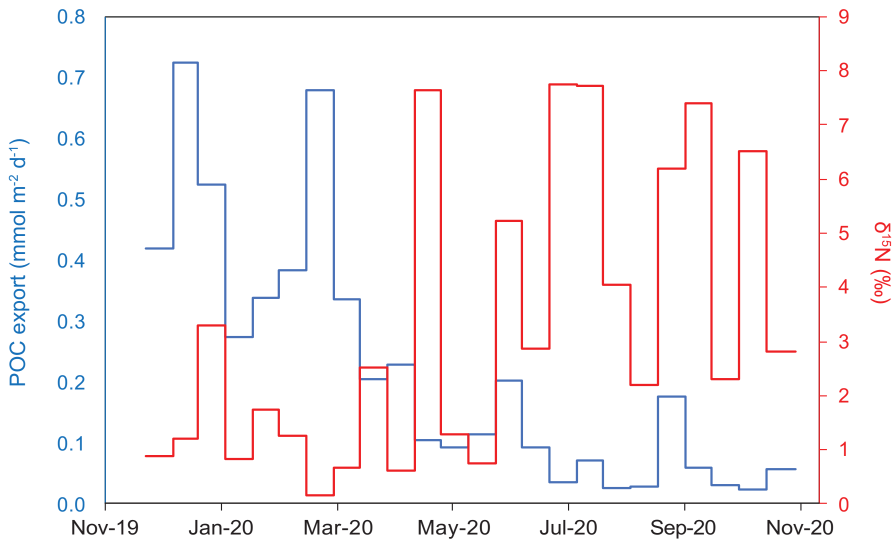


Fig. 4. Seasonal variability of export at 1000 m in the Fe-fertilized patch over 1 year. Fixed mooring deployed at 20°42.12′S; 177°51.96′W from November 2019 to November 2020 (see supplementary materials). The blue line represents POC export fluxes (mmol C m^{-2} day^{-1}), and the red line represents the $\delta^{15}N$ (‰) signature of exported PON.

Table 1. Carbon and Fe budgets in the naturally fertilized region of the Tonga volcanic arc and the distal reference site, as well as comparisons with natural fertilizations in HNLC regions. Dashes indicate that these data are available but were not relevant to this study. n.a., not applicable.

Parameter	TONGA		CROZEX*†	KEOPS*‡
	+Fe (volcano 1)	–Fe (gyre)		
Bloom area (km^2)	360,000	No bloom	90,000	45,000
Bloom duration (days)	180	n.a.	58	75 to 105
Integrated Chla over the euphotic zone (mg Chla m^{-2})	39	n.a.	98.1	72 to 318
Vertical diffusivity (K_z) (m^2 s^{-1})	3.7 ± 1.9 × 10^{-5}	5.2 ± 9.6 × 10^{-6}	–	–
Vertical DFe gradient (mol m^{-4})	3.1 ± 4.7 × 10^{-8}	7.8 ± 3.1 × 10^{-11}	–	–
Vertical DFe diffusive flux (mmol Fe m^{-2} day^{-1})	1.1 ± 1.7 × 10^{-4}	3.5 ± 3.1 × 10^{-8}	6.0 × 10^{-5}	3.1 × 10^{-5}
Atmospheric DFe supply (mmol Fe m^{-2} day^{-1})§	2.0 × 10^{-5}	2.5 × 10^{-5}	1.0 × 10^{-4}	1.7 × 10^{-6}
Horizontal DFe supply (mmol Fe m^{-2} day^{-1})	0¶	0¶	3.9 × 10^{-4}	1.9 × 10^{-4}
Total DFe supply (mmol Fe m^{-2} day^{-1})	1.3 × 10^{-4}	2.5 × 10^{-5}	5.5 × 10^{-4}	2.2 × 10^{-4}
Total annual DFe supply (mmol Fe m^{-2})	4.7 × 10^{-2}	0.9 × 10^{-2}	20.0 × 10^{-2}	8.1 × 10^{-2}
POC export 170 m (mmol C m^{-2} day^{-1})	3.2	1.7	1.3**††	24.5**‡‡
POC export 270 m (mmol C m^{-2} day^{-1})	3.9	1.4		
"Excess" C sequestration efficiency 170 m (mol C mol^{-1} Fe)	13,600	n.a.	8640**	154,000**
"Excess" C sequestration efficiency 270 m (mol C mol^{-1} Fe)	23,000	n.a.		

*See Morris and Charrette (*59*). †See Pollard *et al.* (*39*). ‡See Blain *et al.* (*49*), updated by Chever *et al.* (*60*). §See Guieu *et al.* (*11*). ¶The main flux is from below; lateral advection is likely negligible. **Value for 200 m. ††Interpolated Th-derived POC export flux. ‡‡Th-derived POC export flux.

of the Chla increase after the eruption was distorted by the presence of abundant volcanic particles suspended in the water column after the eruption (*37*).

Looking more deeply into the 25-year monthly Chla time series (Fig. 2B and fig. S10), we find that, despite interannual variability, the bloom develops every year for at least 6 months in austral summer. This seasonal characteristic is probably linked to the thermal fitness of *Trichodesmium*, which are thought to only bloom at temperatures >25°C, reached in the WTSP between November and April (austral summer). This thermal constraint for *Trichodesmium* also probably explains why the bloom does not extend south of ~23°S, which marks the location of the 25°C isotherm. Further south, surface waters are also depleted in nitrate, and in the absence of diazotrophs, the newly emitted Fe supplied by hydrothermal vents along the arc cannot be taken up to produce new biomass. Because models predict a sea surface temperature increase of 1.5°C by the end of the century in the Tonga arc region (*38*), it is possible that the thermal constraint for diazotrophs will be relieved and the bloom will spread further south in the future.

To properly account for the seasonal variability of export, we deployed a moored sediment trap at 1000 m for a full annual cycle in the fertilized patch (see supplementary materials). We show that POC export was five times higher in austral summer than in winter (Fig. 4), resulting in a seasonally integrated (six summer months) POC export of 74 mmol C m^{-2}, that is, 80% of the annual POC export flux. The low δ^{15}N signature of sinking PON during these summer months (~0 to 1‰) confirms that N_2 fixation supports most export production during that season. In austral winter, the higher sinking PON δ^{15}N values (2 to 8‰) suggest that other N sources (i.e., deep nitrate) fuel the low export production. For comparison, the annual POC flux measured here in the subtropical ocean is of the same order of magnitude as that measured in the Southern Ocean in naturally Fe-fertilized waters (*39*, *40*). This suggests that Fe-fertilized regions of the oligotrophic ocean can act as net CO_2 sinks, provided sufficient phosphorus availability. The Fe-fertilized WTSP is a distinctive ecosystem that allows C sequestration supported by N_2 fixation owing to substantial winter phosphate replenishment (not associated with nitrate supply because of the decoupling of phosphacline and nitracline depths) (*25*) and to a turbulent diffusive flux of Fe and phosphate (19 mmol mol^{-1}) that meets *Trichodesmium* requirements (*41*). Furthermore, *Trichodesmium* can reduce its P quotas under P stress (*42*, *43*) and actively uses dissolved organic P compounds in this region (*44*), likely enhanced by high Fe availability (*45*). Taken together, these findings suggest that P and Fe work in concert to trigger extensive diazotroph blooms and C sequestration by N_2-based production in the WTSP.

Compared with shelf-driven natural Fe fertilizations that occur in HNLC (high nutrient, low chlorophyll) waters of the Southern Ocean, the TONGA bloom is generally more temporally and spatially extensive despite its lower intensity (depth-integrated Chla) (Table 1) (*39*, *46*–*48*). The total DFe flux in this study (130 nmol Fe m^{-2} day^{-1}) was generally lower than that measured in Fe-enriched waters downstream of the Kerguelen plateau [KEOPS cruise in 2005, 222 nmol Fe m^{-2} day^{-1} (*49*)] and downstream of the Crozet plateau [CROZEX cruise in 2004, 550 nmol Fe m^{-2} day^{-1} (*39*)] (Table 1). However, unlike HNLC regions, the surface waters of the WTSP are nitrate depleted, and only N_2-fixing organisms can exploit this newly emitted Fe to build biomass and drive carbon export to the deep ocean, as long as sufficient phosphorus remains available. Based on the excess POC export and the excess of DFe supply at the time of the cruise (Table 1), we calculated a C sequestration efficiency (defined as the ratio of the excess POC export to the amount of excess DFe supplied) of 13,600 and 23,000 mol C mol^{-1} Fe (at 170 and 270 m, respectively). Although comparisons between studies need to be considered with caution, given the different methods used and timescales considered, to estimate both excess Fe supply and POC export, this sequestration efficiency is higher than those from artificial mesoscale Fe-addition experiments [e.g., 4300 mol C mol^{-1} Fe for SOFeX (*50*) and 1200 mol C mol^{-1} Fe for SERIES (*51*)] and in the range of values measured in naturally fertilized HNLC regions [8600 mol C mol^{-1} Fe for CROZEX (*39*) to 154,000 mol C mol^{-1} Fe during KEOPS (*49*)]. This confirms that natural Fe fertilizations are more efficient for carbon sequestration than purposeful Fe additions. In addition, comparing such estimates from various natural oceanic settings, Le Moigne *et al.* (*52*) suggested that the apparent variability in C sequestration efficiency may be related to the timescale of Fe delivery (slow delivery being more efficient). Therefore, attention must be paid in future studies to the timescale of delivery of this

newly recognized mode of Fe supply through shallow hydrothermalism, including in temperate and polar ecosystems, where this efficiency could be even more important because of higher macronutrient availability.

Our conceptual view of the ocean Fe cycle has greatly evolved over the past 10 years and highlights the importance of hydrothermal activity on the Fe cycle (*53*). Although model simulations suggest that hydrothermal inputs associated with mid-ocean ridges (>2000 m) contribute 23% of the Fe found in the global ocean water column, that Fe only directly supports 3% of carbon export at 100 m (*54*). This is mostly because a large part of that Fe remains in the deep ocean over long timescales (*53*) and needs to be entrained in surface waters before potentially affecting photosynthetic communities (*46*, *48*). Fe from intermediate-depth (~1000 m) vents may also be transported long distances and, with the condition that these waters upwell, influence marine ecosystems located thousands of kilometers from the site of discharge (*55*). However, hydrothermal venting also occurs at shallower depths (<500 m) in island arc systems such as the Tonga arc. Even if scavenging and precipitation removes part of this newly emitted Fe from the dissolved pool (*23*), such shallow sources can rapidly supply Fe to overlying surface photosynthetic communities compared with Fe emitted at greater depths (*23*, *55*). In the oligotrophic ocean, the implications of such shallow hydrothermal Fe fertilization are highly important because they directly fuel surface diazotrophs and the export of organic matter to the deep ocean, representing regional hotspot sinks of atmospheric CO_2. We demonstrate here that shallow hydrothermal sources also represent a triggering factor on diazotroph blooms in regions where the atmospheric supply of DFe is virtually absent. Such forcing is of the utmost importance to study because climate models predict an expansion of the oligotrophic gyres (40% of our oceans) (*56*) where diazotrophs will likely thrive. Furthermore, in a warmer, more stratified ocean, such shallow Fe sources are likely to deliver Fe to surface communities more readily than deep sources (*55*). Beyond the oligotrophic oceans, shallow hydrothermal fertilizations are likely to be common in the global ocean because of the high number of shallow hydrothermal vents associated with island arc systems and submarine volcanic calderas (*57*), the exact numbers and locations of which are yet to be discovered (*14*). Such systems are also present at higher latitudes, notably in the HNLC waters in the subarctic Pacific and the Southern Ocean (*57*). Therefore, a comprehensive evaluation of their impact in these severely Fe-limited systems where surface mixed layers reach the intermediate or even the deep ocean water masses is clearly needed. Lastly, the extent to which such hydrothermal-driven biological carbon pump enhancement may have changed atmospheric CO_2 in the past remains unclear. Future studies would be relevant because the hydrothermal flux of Fe has been relatively constant over millennial timescales (*58*).

REFERENCES AND NOTES

1. C. M. Moore *et al.*, *Nat. Geosci.* **6**, 701–710 (2013).
2. N. Gruber, in *Nitrogen in the Marine Environment* (Academic Press, 2008), pp. 1–50.
3. S. Bonnet *et al.*, *ISME J.* **17**, 47–58 (2023).
4. D. M. Karl, M. J. Church, J. E. Dore, R. M. Letelier, C. Mahaffey, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1842–1849 (2012).
5. A. Subramaniam *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10460–10465 (2008).
6. J. P. Zehr, D. G. Capone, *Science* **368**, eaay9514 (2020).
7. C. Lory *et al.*, *ISME Commun.* **2**, 41 (2022).
8. J. A. Sohm, E. A. Webb, D. G. Capone, *Nat. Rev. Microbiol.* **9**, 499–508 (2011).
9. S. Bonnet, M. Caffin, H. Berthelot, T. Moutin, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2800–E2801 (2017).
10. C. Schlosser *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1438–1442 (2014).
11. C. Guieu *et al.*, *Sci. Rep.* **8**, 9075 (2018).
12. L. Bopp *et al.*, *Biogeosciences* **19**, 4267–4285 (2022).
13. C. E. J. de Ronde, G. J. Massoth, E. T. Baker, J. E. Lupton, in *Volcanic, Geothermal, and Ore-Forming Fluids: Rulers and Witnesses of Processes Within the Earth*, vol. 10, S. F. Simmons, I. Graham, Eds. (Society of Economic Geologists, 2005), pp. 91–110.
14. C. R. German *et al.*, *Philos. Trans. R. Soc. Londong Ser. A* **374**, 20160035 (2016).
15. J. E. Lupton, D. G. Pyle, W. J. Jenkins, R. Greene, L. Evans, *Geochem. Geophys. Geosyst.* **5**, Q01003 (2004).
16. P. Stoffers *et al.*, *Geology* **34**, 453–456 (2006).
17. G. Massoth *et al.*, *Geochem. Geophys. Geosyst.* **8**, Q11008 (2007).
18. C. Guieu, S. Bonnet, TONGA GEOTRACES GPpr14 ocean expedition. (2019), https://doi.org/10.17600/18000884.
19. C. Boulart *et al.*, *Commun. Earth Environ.* **3**, 64 (2022).
20. W. W. Chadwick Jr. *et al.*, *Geochem. Geophys. Geosyst.* **15**, 4325–4342 (2014).
21. N. Feuillet *et al.*, *Nat. Geosci.* **14**, 787–795 (2021).
22. J. A. Resing *et al.*, *Nature* **523**, 200–203 (2015).
23. C. Tilliette *et al.*, *Global Biogeochem. Cycles* **36**, e2022GB007363 (2022).
24. J. Lavelle, I. Lozovatsky, D. Smith IV, *Geophys. Res. Lett.* **31**, L10308 (2004).
25. T. Moutin *et al.*, *Biogeosciences* **15**, 2961–2989 (2018).
26. Y.-W. Luo *et al.*, *Earth Syst. Sci. Data* **4**, 47–73 (2012).
27. A. Filella *et al.*, *Front. Mar. Sci.* **9**, 923765 (2022).
28. F. Van Wambeke *et al.*, *Biogeosciences* **15**, 2669–2689 (2018).
29. M. Benavides *et al.*, *ISME J.* **16**, 2398–2405 (2022).
30. A. N. Knapp *et al.*, *Biogeosciences* **15**, 2619–2628 (2018).
31. L. Rousselet *et al.*, *Biogeosciences* **15**, 2411–2431 (2018).
32. A. de Verneil, L. Rousselet, A. M. Doglioli, A. A. Petrenko, T. Moutin, *Biogeosciences* **14**, 3471–3486 (2017).
33. S. E. Beaulieu, E. T. Baker, C. R. German, A. Maffei, *Geochem. Geophys. Geosyst.* **14**, 4892–4905 (2013).
34. C. R. German *et al.*, *Philos. Trans. R. Soc. London Ser. A* **374**, 20160035 (2016).
35. M. Messié, A. Petrenko, A. M. Doglioli, E. Martinez, S. Alvain, *Nat. Geosci.* **15**, 469–474 (2022).
36. B. Barone, R. M. Letelier, K. H. Rubin, D. M. Karl, *Geophys. Res. Lett.* **49**, e2022GL099293 (2022).
37. A. Whiteside *et al.*, *Front. Mar. Sci.* **9**, 1028022 (2023).
38. L. Dhage, M. J. Widlansky, *Earths Futur.* **10**, e2021EF002524 (2022).
39. R. T. Pollard *et al.*, *Nature* **457**, 577–580 (2009).
40. M. Rembauville, I. Salter, N. Leblond, A. Gueneugues, S. Blain, *Biogeosciences* **12**, 3153–3170 (2015).
41. J. Nuester, S. Vogt, M. Newville, A. B. Kustka, B. S. Twining, *Front. Microbiol.* **3**, 150 (2012).
42. D. M. Karl, R. Letelier, D. V. Hebel, D. F. Bird, C. D. Winn, in *Marine Pelagic Cyanobacteria: Trichodesmiurn and Other Diazotrophs*, Nato Science Series C, E. J. Carpenter, D. G. Capone, J. G. Rueter, Eds. (Kluwer Academic Publishers, 1992), pp. 219–237.
43. A. E. White, Y. H. Spitz, D. M. Karl, R. M. Letelier, *Limnol. Oceanogr.* **51**, 1777–1790 (2006).
44. Z. Liang, R. T. Letscher, A. N. Knapp, *Nat. Geosci.* **15**, 651–657 (2022).
45. T. J. Browning *et al.*, *Nat. Commun.* **8**, 15465 (2017).
46. M. Ardyna *et al.*, *Nat. Commun.* **10**, 2451 (2019).
47. S. Blain, S. Bonnet, C. Guieu, *Biogeosciences* **5**, 269–280 (2008).
48. C. M. S. Schine *et al.*, *Nat. Commun.* **12**, 1211 (2021).
49. S. Blain *et al.*, *Nature* **446**, 1070–1074 (2007).
50. K. O. Buesseler, J. E. Andrews, S. M. Pike, M. A. Charette, *Science* **304**, 414–417 (2004).
51. P. W. Boyd *et al.*, *Nature* **428**, 549–553 (2004).
52. F. A. C. Le Moigne *et al.*, *Geophys. Res. Lett.* **41**, 4619–4627 (2014).
53. A. Tagliabue *et al.*, *Nature* **543**, 51–59 (2017).
54. A. Tagliabue, O. Aumont, L. Bopp, *Geophys. Res. Lett.* **41**, 920–926 (2014).
55. W. Jenkins *et al.*, *Earth Planet. Sci. Lett.* **539**, 116223 (2020).
56. J. J. Polovina, E. A. Howell, M. Abecassis, *Geophys. Res. Lett.* **35**, L03618 (2008).
57. J. A. Hawkes, D. P. Connelly, M. J. A. Rijkenberg, E. P. Achterberg, *Geophys. Res. Lett.* **41**, 942–947 (2014).
58. N.-C. Chu *et al.*, *Earth Planet. Sci. Lett.* **245**, 202–217 (2006).
59. P. J. Morris, M. A. Charette, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **90**, 147–157 (2013).
60. F. Chever, G. Sarthou, E. Bucciarelli, S. Blain, A. R. Bowie, *Biogeosciences* **7**, 455–468 (2010).

ACKNOWLEDGMENTS

The dataset presented here resulted from the efforts of many individuals both onboard and on land who contributed to the success of the expedition. We thank the captain and the crew of the reserach vessel *L'Atalante* (TGIR Flotte, operated by IFREMER), in particular, S. Laville Saint-Martin and J. Le Doare for their help during the multibeam survey, as well as the Technical Division INSU for technical support and equipment provision. We thank the captain and the crew of the research vessel *Alis* (TGIR Flotte, operated by IFREMER) for the safe recovery of the fixed mooring in particularly difficult weather conditions. We thank C. Dimier and the SAPIGH platform of the Institut de la Mer de Villefranche (IMEV) for sampling on board and performing the pigment analyses, and M. Tantillo for help with the noble gas extraction from water samples and isotope analysis performed at the noble gas laboratory of INGV-Palermo. We also thank H. Lucas Jr. for English editing. We thank J. P. Gattuso, S. Blain, A. Tagliabue, and L. Guidi for their advice on the manuscript. Lastly, we thank J. Resing and an anonymous reviewer for their constructive and valuable comments during the review process. **Funding:** This work was funded by Agence Nationale de Recherche grant ANR-18-CE01-0016 (S.B. and C.G.), A-MIDEX grant TONGA (S.B. and C.G., Excellence Initiative of Aix-Marseille University, a French "Investissements d'Avenir" program). Institut National des Sciences de l'Univers Les Enveloppes Fluides et l'Environnement grant TONGA (S.B. and C.G.), TGIR Flotte océanographique française TONGA cruise (C.G. and S.B.), the Institut de recherche pour le Développement (IRD) (S.B.), a MEMESTRA grant (C.B.), and National Science Foundation grant NSF-OCE 1829797 (A.N.K.). Finally, this project has received funding from the European FEDER Fund under project 1166-39417. **Author contributions:** Conceptualization: S.B., C.G.; Methodology: S.B., C.G., C.B., F.G.; Investigation: S.B., C.G., V.T., C.B., F.G., M.B., D.G.-S., J.-M.G., O.G., J.H., C.L., E.P.-V., G.S., C.T.; Data curation: S.B., C.G., V.T., C.B., P.B.-A., C.S., F.G., M.B., A.N.K., Y.C., D.G.-S., H.J.F., J.H., N.L., S.N., A.L.R., G.S., C.T., A.L.-D., F.A.C.L.M., M.J.-C.; Visualization: S.B., C.G., C.B., F.G., P.B.-A., A.N.K., and H.J.F., A.L.-D., F.A.C.L.M., M.J.-C.; Writing – original draft: S.B., C.G.; Writing – review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The $\delta^{15}N$ data are deposited in the BCO-DMO database (https://www.bco-dmo.org/dataset/869963), the data are deposited in the SEANOE database (https://campagnes.flotteoceanographique.fr/campagnes/18000884/).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq4654
Materials and Methods
Figs. S1 to S10
Tables S1 to S8
References (*61*–*96*)

Submitted 12 April 2022; accepted 21 April 2023
10.1126/science.abq4654

CELL BIOLOGY

Ectocytosis renders T cell receptor signaling self-limiting at the immune synapse

Jane C. Stinchcombe[1], Yukako Asano[1], Christopher J. G. Kaufman[1], Kristin Böhlig[2], Christopher J. Peddie[3], Lucy M. Collinson[3], André Nadler[2], Gillian M. Griffiths[1]*

Cytotoxic T lymphocytes (CTLs) kill virus-infected and cancer cells through T cell receptor (TCR) recognition. How CTLs terminate signaling and disengage to allow serial killing has remained a mystery. TCR activation triggers membrane specialization within the immune synapse, including the production of diacylglycerol (DAG), a lipid that can induce negative membrane curvature. We found that activated TCRs were shed into DAG-enriched ectosomes at the immune synapse rather than internalized through endocytosis, suggesting that DAG may contribute to the outward budding required for ectocytosis. Budding ectosomes were endocytosed directly by target cells, thereby terminating TCR signaling and simultaneously disengaging the CTL from the target cell to allow serial killing. Thus, ectocytosis renders TCR signaling self-limiting.

The ability to kill multiple target cells in rapid succession is an important factor in the effectiveness of cytotoxic T lymphocytes (CTLs), both in the immune response and in cancer immunotherapies. Such serial killing requires a carefully controlled transition from recognition to detachment from each target cell encountered. The TCR-CD3 complex is formed by TCRs (α and β) that recognize peptide–major histocompatibility complex I (pMHCI) and a cluster of CD3 chains (γ, δ, ε, and ζ) required for signaling (fig. S1A). TCR engagement with pMHCI on target cells triggers rapid phosphorylation of the TCR complex through lymphocyte-specific protein tyrosine kinase (Lck) and Zeta-chain-associated protein kinase 70 (Zap70), which initiates downstream signaling [reviewed in (*1*)]. TCR activation can down-regulate surface TCR expression through endocytosis [reviewed in (*2*)], and studies using artificial immunological synapses on lipid bilayers have also shown that TCRs can be shed in exosomes (*3*) and right-side-out microvesicles (ectosomes) (*4*, *5*). What determines the route of TCR down-regulation is not known.

Tracking the TCR in the immune synapse

To understand how the seamless transitions from TCR recognition to signaling and detachment might be linked, we tracked the fate of TCRs at different stages of immune synapse formation with high-resolution three-dimensional (3D) imaging. We labeled CD3ζ with engineered pea ascorbate peroxidase (APEX) (*6*) (fig. S1A) and imaged TCR localization with electron microscope (EM) tomography. CD3ζ-APEX colocalized with endogenous CD3ε in CTLs (fig. S1B). CTLs engaged targets at different times after mixing, allowing us to capture different stages of

[1]Cambridge Institute for Medical Research, Keith Peters Building, Cambridge CB2 0XY, UK. [2]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [3]The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK.
*Corresponding author. Email: gg305@cam.ac.uk

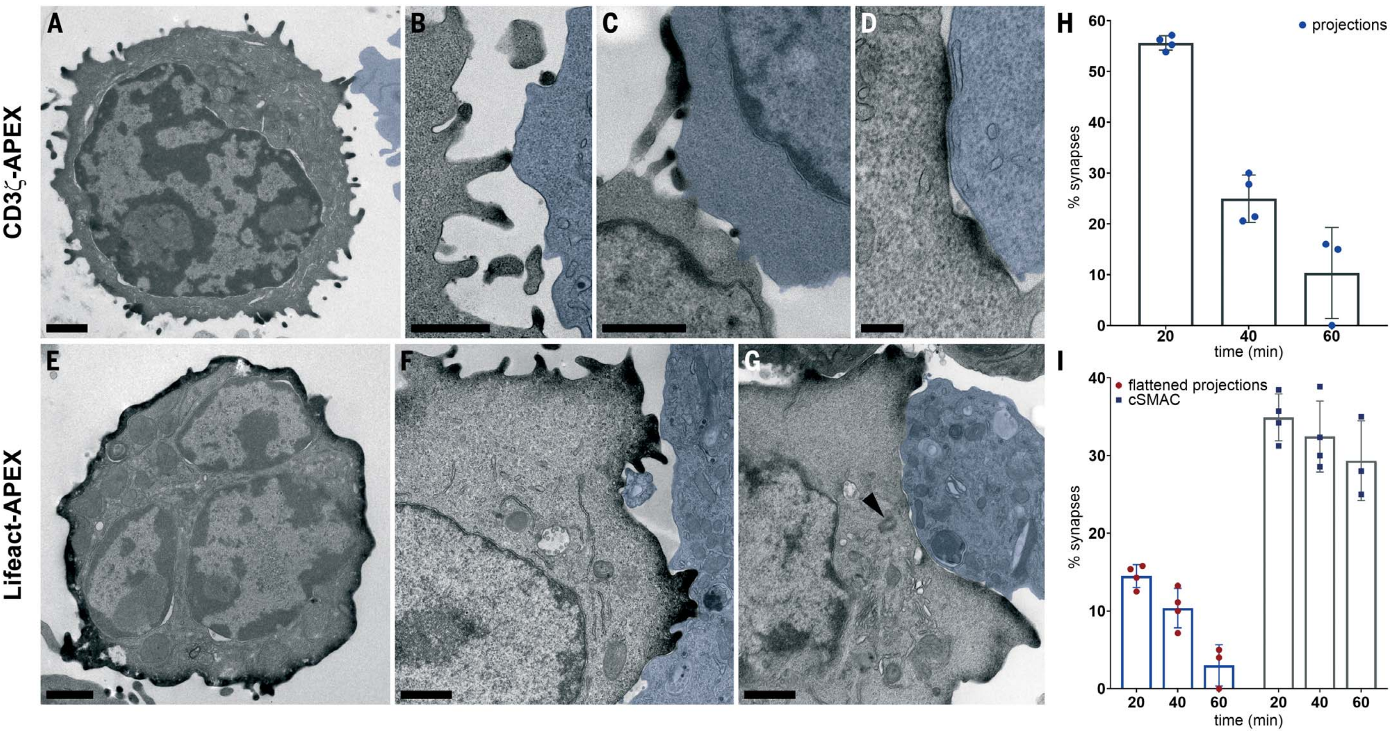


Fig. 1. CD3ζ-APEX and Lifeact-APEX localization in CTLs. (**A** to **D**) TEM images of CTLs expressing CD3ζ-APEX incubated with target cells (blue) for 20 min, showing [(A) and (B)] projections, (C) flattened projections, and (D) cSMAC. (**E** to **G**) TEM images of CTL expressing Lifeact-APEX with cells in (F) and (G) incubated with target cells (blue) for (F) 40 min and (G) 20 min; black arrowhead marks centrosome. Sections in (A) to (G) are 50 to 70 nm except (C), which is 250 nm. Scale bars, 1 μm [(A) to (C) and (E) to (G)] and 200 nm (D). Images are representative of (A) 38, (B) 42, (C) 54, and (D) 67 examples from four independent experiments and (E) 54, (F) 95, and (G) 28 examples from five independent experiments. (**H** and **I**) Percentage of immune synapse images from CTLs and targets incubated for 20, 40, and 60 min showing CD3ζ-APEX in (H) projections (blue circles) and (I) flattened projections (red circles) or cSMAC (blue squares) from 107 (20 min), 146 (40 min), and 49 (60 min) APEX-labeled immune synapses. Error bars show mean values ± SD from four independent experiments.

immune synapse formation up to the point of fixation in our samples. CD3ζ-APEX, identified by its electron-dense reaction product, labeled the cell surface, including the tips of lamellipodial projections (fig. S2, A and B) that make first contact with target cells (Fig. 1, A and B) (*7*). These projections flattened against the target (Fig. 1C) and were lost over time as the two cells formed extended areas of close membrane contact that were lined with CD3ζ-APEX (Fig. 1, D, H, and I).

Lifeact-APEX, which binds to F-actin, was associated with the cortex of isolated CTLs (Fig. 1E) and in lamellipodial projections contacting targets (Fig. 1F) but was depleted where the immune synapse formed an area of tight membrane contact opposite the site of centrosome polarization (Fig. 1G). We also observed this result in 3D when using focused ion beam scanning EM (FIB-SEM) (fig. S2, C to E, and movie S1). Thus, TCRs in actin-rich lamellipodia make the first contact between CTLs and targets, with the lamellipodia rapidly flattening to form close membrane contacts lined with TCRs, and depleted of actin as the synapse forms.

Ectocytosis rather than endocytosis at the immune synapse

We examined the 3D localization of CD3ζ-APEX in 12 separate tomograms, each spanning ~1 μm across the immune synapse, using high-resolution EM tomography (Fig. 2, A and B, and movie S2). Modeling of the reconstructed tomograms (movie S3) showed CD3ζ-APEX (green) extending across an area up to 2 μm in length and >1 μm in width on the CTL membrane (magenta), corresponding to TCR clusters observed with light microcopy that have been termed "central supramolecular activation clusters (cSMACs)" (*8*), where TCRs and pMHCI are engaged and CTLs and target membranes are tightly apposed. Although we could find CD3ζ-APEX associated with clathrin-coated invaginations and tubules internalizing into the CTL, most of CD3ζ-APEX was present in 37- to 210-nm-diameter ectosomes (*9*) budding outward from the edges of the CD3ζ-APEX–labeled cSMAC (Fig. 2, A to C, and movies S2 and S3). Quantitation revealed 98% of vesicular CD3ζ in ectosomes (Fig. 2D), suggesting that TCR is down-regulated through ectocytosis rather than endocytosis across the immune synapse.

Functions of ectocytosis

Three-dimensional views and models of tomograms with immune synapses enriched in CD3ζ-ectosomes (Fig. 3, A to F, and movies S4 to S7) revealed multiple smaller regions of membrane-associated CD3ζ (42 nm by 250 nm to 800 nm by 250 nm) with ectosomes either budding from the periphery of cSMACs (Fig. 3B) or detached from the CTL membrane (Fig. 3, A, C, E, and F). Quantitation demonstrated that membrane-associated CD3ζ decreased as ectosomes increased (Fig. 3G) and that the number of conjugates with CD3ζ-ectosomes increased over time (Fig. 3H), suggesting that the reduction in area of TCRs across the membrane resulted from ectocytosis. Both 3D tomograms and conventional EM also revealed that the tight membrane contacts between CTLs and targets were lost where ectosomes budded from the CTL, creating localized cell separation between CTLs and targets (Figs. 3A and 4A). Quantitation with light microscopy showed that ectocytosis increased with TCR signal strength whereas CTL surface–TCR expression decreased (Fig. 4B and fig. S3). TCR signal strength also increased detachment of CTLs from targets (Fig. 4C). These results suggest that ectocytosis both removed TCR from the immune synapse and initiated detachment between CTLs and targets as TCR ectosomes were shed.

Fate of shed ectosomes

Our EM tomograms also revealed the fate of shed ectosomes. We found that CD3ζ-ectosomes were closely associated with the target cell membrane, with 83% of budding ectosomes ($n = 29$) and 71% of budded ectosomes ($n = 293$) in contact with the target membrane when ectosomes separated from the CTL (Figs. 3A and 4A, and movie S5). Using CTL-target conjugates, we found that ectosomes were taken up into target cells with invaginations containing CD3ζ-ectosomes that reached deep into

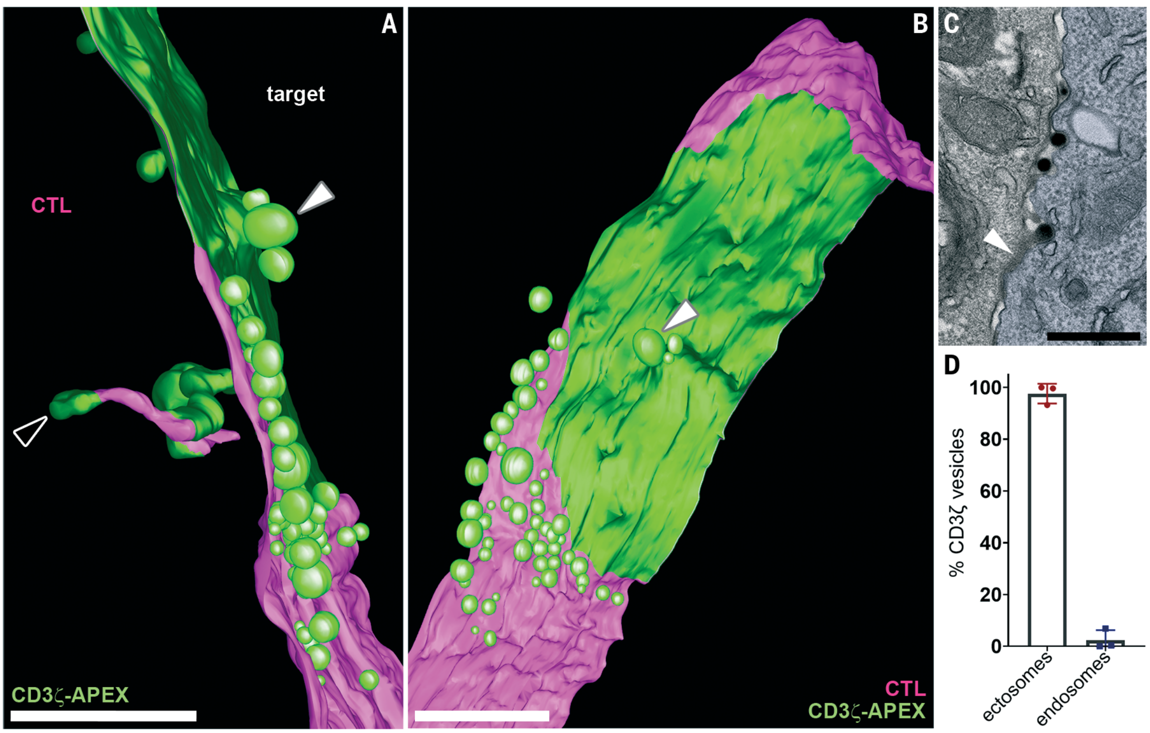


Fig. 2. TCR is down-regulated through ectocytosis rather than endocytosis across the synapse. (**A** and **B**) Images from the 3D model generated from a tomogram reconstruction through ~1 μm across the immune synapse (movies S2 and S3), showing the CTL membrane (magenta) and CD3ζ-APEX labeling (green) with (A) side and (B) en face views. The white arrowheads in (A) and (B) identify the same ectosome; the white-edged arrowhead identifies a tubular endosome. Images are representative of six independent tomograms. Scale bars, 500 nm. (**C**) TEM image of a CTL-target conjugate, showing CD3ζ-APEX in protrusions and buds emerging from the edges of membrane-clustered CD3ζ (white arrowhead) after incubation with target cells (blue) for 40 min. Image is representative of 61 examples from five independent experiments. Scale bar, 500 nm. (**D**) Percentage of CD3ζ-APEX–labeled endosomes versus ectosomes at the immune synapse. Error bars show mean values ± SD from three independent tomograms.

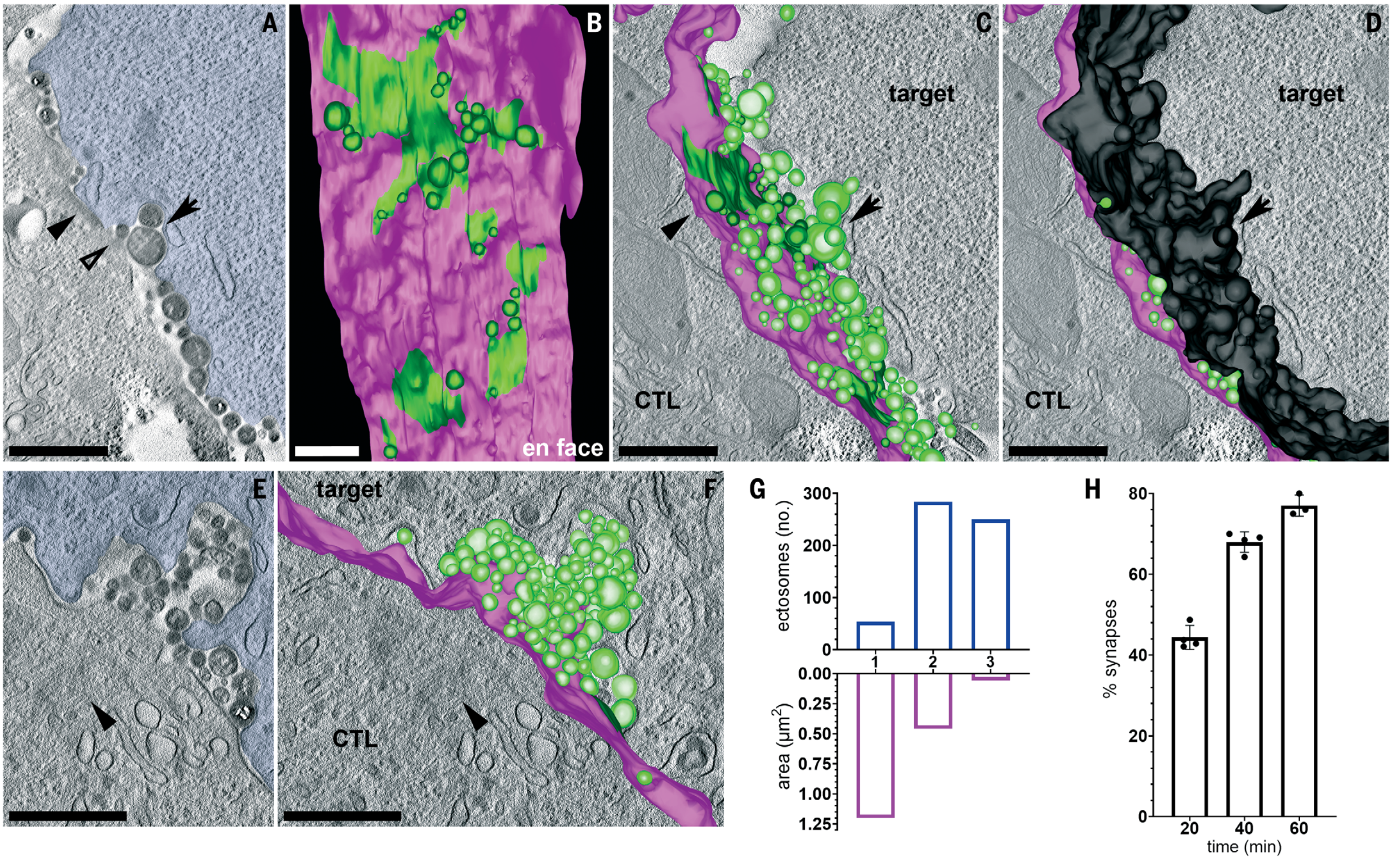


Fig. 3. cSMACs diminish through ectocytosis. (**A** to **D**) Images from a tomogram reconstruction through ~1.25 μm across the immune synapse (movie S4) and associated 3D model (movie S5) with CTL membrane (magenta), CD3ζ-APEX (green), and targets false colored blue or target membrane modeled in black. (A) The image shows a single plane from the tomogram series. (B) The en face 3D model view shows CD3ζ-APEX plasma membrane labeling plus ectosomes budding from the CTL surface. The images in (C) and (D) show single-tomogram planes with the 3D model superimposed, including (D) the target membrane. The black arrowheads in (A) and (C) point to the same area of CD3ζ-APEX–labeled CTL membrane; the arrows in (A), (C), and (D) point to the same target cell invagination. The open arrowhead in (A) shows a budding ectosome. (**E** and **F**) Single images from a tomogram series through ~1 μm across the immune synapse (movie S6) with (E) the target cell false colored blue and (F) the 3D model (movie S7) superimposed showing the CTL membrane (magenta) and CD3ζ-APEX labeling (green). The black arrowheads indicate the polarized centrosome. Images in (A) to (F) are representative of six independent tomograms. Scale bars, 500 nm. (**G**) Comparison of total surface area of membrane labeled with CD3ζ-APEX (magenta) with the number of CD3ζ-APEX–labeled ectosomes (blue) across each tomogram. Data for (1), (2), and (3) are from movies S3, S5, and S7, respectively. (**H**) Percentage of immune synapses with CD3ζ-APEX within ectosomes at different times after cell mixing. Data in (G) and (H) areas described in Fig. 1, H and I, and are from 107 (20 min), 146 (40 min), and 49 (60 min) APEX-labeled immune synapses. Error bars show mean values ± SD from four independent experiments.

the target cell (Fig. 4, D to F, and figs. S4 and S5). These ectosomes were often decorated with clathrin coats, sometimes with endoplasmic reticulum (ER) closely associated (Fig. 4E). CD3ζ-ectosomes were also seen in vacuolated targets that appeared to be dying (Fig. 4F). We observed similar results with untransfected CTLs (Fig. 4G). Thus, budding ectosomes were taken up into dying targets in clathrin-decorated structures.

To understand whether ectocytosis effectively down-regulated signaling from shed receptors, we asked whether other proteins associated with TCR activation were contained in ectosomes. Using confocal and structured illumination microscopy (SIM), we imaged ectosomes shed from untransfected CTLs. Endogenous CD3ζ, CD3ε, TCRβ, CD8, and Zap70 were all detected in ectosomes that transferred to target cells across the synapse (Fig. 5, A and D, and figs. S4 to S6). By contrast, Lck, the kinase that initiates TCR activation, remained clustered on the CTL membrane at the synapse (Fig. 5, A and D, and fig. S6). We next asked whether activated TCRs were shed into ectosomes by using antibodies against their phosphorylated forms. Activated pCD3ζ and pZap70 were localized in ectosomes (Fig. 5, A and E, and fig. S6). There was no evidence of Zap70 in intracellular vesicles within the CTL (fig. S7). Thus, released ectosomes contained activated TCRs and Zap70 but not the initial activating kinase Lck, preventing further activation and supporting the idea that ectocytosis can act as a mechanism to terminate signaling from activated TCRs that are shed.

We asked whether ectocytosis of TCRs was linked to the exocytosis of cytolytic granules that is triggered by TCR activation. Rab27-deficient CTLs are unable to release their cytolytic granules, fail to kill target cells, and, like other granule-deficient CTLs, continue to recognize targets over prolonged periods (*10*, *11*). Nevertheless, we found that conjugates formed by Rab27-deficient CTLs generated ectosomes that were transferred across the synapse to target cells (fig. S8). We also confirmed that the cytolytic granule protein granzyme B (GrB) did not colocalize with CD3ε within CTL and was not found in ectosomes (Fig. 5D and fig. S4). Thus, ectocytosis is independent of cytolytic granule ectocytosis, and ectosomes do not contain cytolytic proteins.

The role of DAG in ectocytosis

We have previously shown that TCR activation creates an area of membrane lipid and

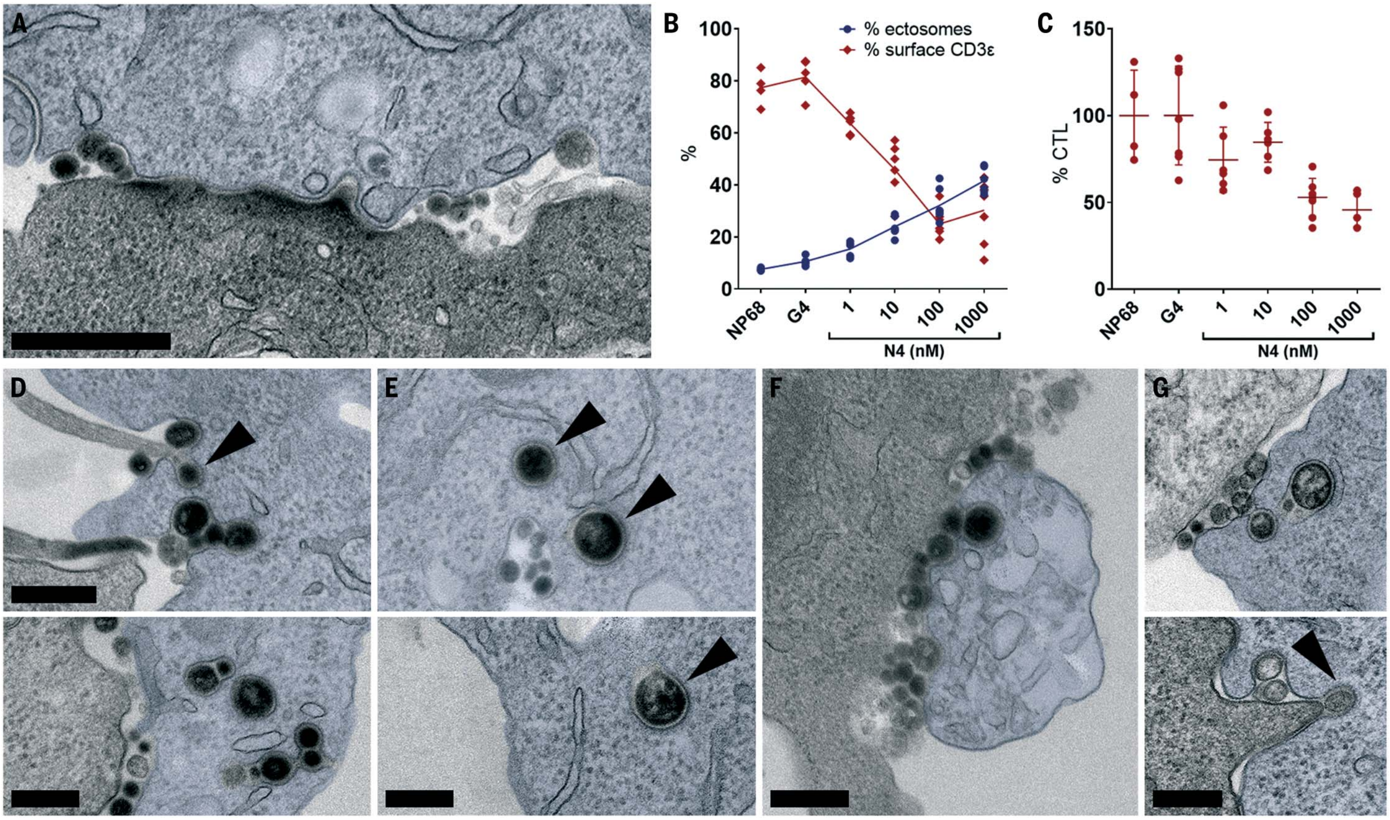


Fig. 4. Ectosomes are internalized into target cells through clathrin-mediated endocytosis. (**A**) TEM image through an immune synapse with target cell in blue, showing CD3ζ-APEX ectosomes budding at the edges of membrane-clustered CD3ζ (cSMAC) and causing separation of CTL and target membranes. The image is representative of 61 examples from five independent experiments. Scale bar, 500 nm. (**B**) Percentage of CTLs with CD3ε surface labeling (red graph) and percentage of target cells decorated with transferred CD3ε-labeled ectosomes (blue graph), analyzed as shown in fig. S3, from CTL-target conjugates stimulated with irrelevant (NP68), weak (G4), or strong (N4) peptides. Each data point shows quantitation per 40× field from 163 to 358 CTLs and 274 to 463 targets per condition; error bars show mean value ± SD from four to five technical replicates; and data are representative of three independent experiments. (**C**) CTL detachment from CTL-target conjugates over 35 min, expressed as percentage of CTLs in NP68 sample (51 CTLs per image field). Datasets are as described in (B). (**D** to **F**) TEM images of CD3ζ-APEX expressing CTL, showing CD3ζ-ectosomes internalized into target cells (blue), including (F) dying cells with vacuolated ER. The black arrowheads indicate clathrin coats. (**G**) TEM images of immune synapses between untransfected CTL and targets (blue), showing ectosomes internalized into endocytic vesicles with clathrin coats (black arrowhead). Images in (D) to (G) are from 50- to 70-nm sections of CTL-target conjugates fixed after 40-min or (F) 60-min coincubation and are representative of [(D) and (E)] 91 and (F) 28 synapses from four independent experiments and of (G) >50 synapses from seven independent experiments. Scale bars, 250 nm.

phosphoinositide specialization within the immune synapse upon phospholipase C–γ (PLC-γ) cleavage of phosphatidylinositol 4,5-bisphosphate (PIP_2), generating DAG (*12*) (fig. S1A). Studies on neutrophils, in which the term "ectocytosis" was first used to define the generation of right-side-out plasma membrane vesicles, demonstrated the preferential sorting of both DAG and specific proteins into shed ectosomes (*9*). Furthermore, accumulation of DAG in erythrocytes was proposed to create negative membrane curvature, forcing outward budding of vesicles from the plasma membrane (*13*, *14*). We therefore asked whether DAG might contribute to ectocytosis of TCRs from CTLs. To visualize DAG in both fixed and live CTL-target conjugates, we used protein kinase C–γ (PKCγ) or PKCε bioprobes that bind DAG (*12*, *15*, *16*) (fig. S9 and movie S8). We found that DAG appeared across the immune synapse and separated into ectosomes that colabeled with anti-TCR antibodies (Fig. 5B). Using live-cell imaging, we were able to image DAG-enriched ectosomes emerging from the synapse as DAG was generated during target cell recognition (Fig. 5C and movie S9). Thus, DAG is a membrane constituent of TCR-ectosomes and may contribute to the negative membrane curvature required to initiate ectocytosis of TCRs from CTLs.

Our findings support a model in which the negative membrane curvature induced by the generation of DAG during TCR signaling could initiate ectocytosis of activated TCRs (fig. S10). We propose that the production of DAG around activated TCRs would contribute to the negative membrane curvature required to initiate ectocytosis, linking TCR activation to the shedding of activated TCRs. Ectosome budding also initiates localized detachment between CTL and target within the immune synapse, allowing CTLs to detach from targets once no additional TCRs are recruited and target recognition has ceased.

This model of DAG-driven ectocytosis does not preclude a potential role for vacuolar protein sorting 4 (Vps4) in ectosome scission (*4*). However, our experiments showed that expression of dominant negative Vps4 caused TCR to be retained within enlarged endosomal compartments in the CTL, thereby preventing synapse formation and Lck recruitment (fig. S11). Consequently, a potential role for Vps4 in ectocytosis cannot be resolved. However, our findings are consistent with potential roles for actin in cell separation (*17*, *18*) because actin is depleted and restored in tune with TCR activation (*12*).

Discussion

Using EM tomography, we generated 3D views over >1-μm depths across the immune synapse

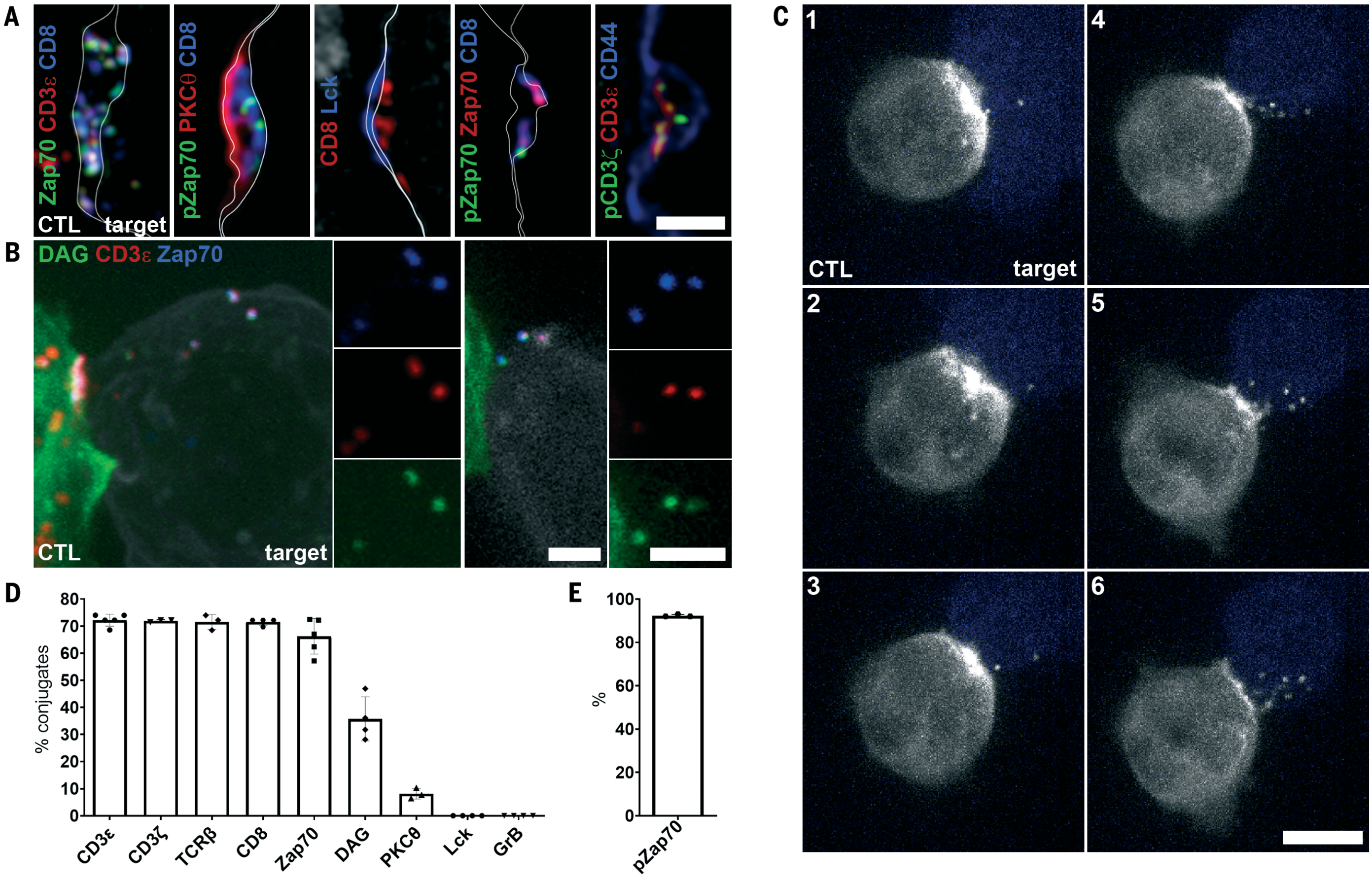


Fig. 5. Activated TCRs and Zap70 are shed into ectosomes. (**A**) SIM imaging showing the central area of immune synapses between untransfected CTLs and target cells labeled with antibodies as indicated. Each panel shows a single-section image with the CTL on the left and target on the right, and cell membranes either traced in white or delineated with CD44 labeling (blue). Images are representative of (left to right) 72, 17, 98, 45, and 25 synapses from three to five independent experiments. Scale bar, 1 μm. (**B**) (Left) Projected *z*-stack and (right) single confocal images through synapses between CTL expressing enhanced green fluorescent protein (EGFP)–tagged bioprobes for DAG (left, PKCε; right, PKCγ-C1) (green) and targets (white), and labeled for CD3ε (red) and Zap70 (blue). Images are representative of 29 (PKCε) and 27 (PKCγ-C1) synapses across two and four independent experiments, respectively. Scale bars, 2 μm. (**C**) Image series from live imaging of CTLs expressing PKCε-EGFP bioprobe for DAG (white) synapsed with target cells (blue) (movie S9). Images are projections of 33 *z*-planes taken at 68-s intervals. Data are representative of six movies from four independent experiments. Scale bar, 5 μm. (**D**) Percentage of conjugates with ectosomes transferred to target cells after 35- to 45-min incubation with CTLs, analyzed with confocal microscopy for each marker as shown. The EGFP-tagged DAG bioprobe signal will diminish in acidic endocytic compartments. (**E**) Percentage of Zap70 or CD8 ectosomes also labeled with phosphorylated Zap70 (pZap70). Each data point in (D) and (E) represents a separate biological repeat (three to five per marker) and graphs include data from more than 1034 (CD3ε), 158 (CD3ζ), 152 (TCRβ), 180 (CD8), 492 (Zap70), 144 (DAG), 374 (PKCθ), 289 (Lck), 114 (GrB), and 57 (pZap70) synapses labeled as indicated. Additional representative confocal images used in quantitation are provided in figs. S3, S5, and S6. Error bars show mean values ± SD from three to five independent experiments.

formed between CTLs and targets. Together with >800 transmission EM (TEM) images and >8000 immunofluorescence images, we examined the fate of the TCR within the immune synapse. We found that activated TCRs were shed into DAG-enriched ectosomes at the immune synapse rather than internalized through endocytosis. Ectosomes remained tightly bound by target cells during budding and were internalized into clathrin-decorated structures within the targets. As ectosomes budded from the CTLs, they created an area of cell separation that allowed CTLs to detach from their targets as TCR signaling ceased. Both ectocytosis and CTL detachment increased with signal strength.

Our results provide insights into the well-established phenomenon of TCR down-regulation after activation (*19*–*22*). Although TCR endocytosis and subsequent degradation has been considered the key mechanism for TCR down-regulation (*2*, *19*–*23*), we now show that this is not so at the immune synapse, where activated TCRs are preferentially shed through ectocytosis into DAG-enriched ectosomes. Two lines of evidence support our observation that endocytosis of activated TCR is inhibited across the synapse. First, structural studies have demonstrated that adapter protein-2 (AP2) needs PIP_2 in the membrane to recruit clathrin and initiate endocytosis (*24*). Thus, after TCR activation when PIP_2 is depleted across the immune synapse (*12*), AP2 will be unable to initiate endocytosis. Second, it has been shown that AP2 cannot bind YXXΦ endocytosis motifs (where Y is tyrosine, X is any amino acid, and Φ is any bulky hydrophobic amino acid) when the tyrosine is phosphorylated (*25*, *26*). Because the 20 functional YXXΦ endocytosis motifs are embedded in immunoreceptor tyrosine-based activation motifs (ITAMs) in which the tyrosine residues are phosphorylated during TCR activation (*1*, *27*), activated TCRs will not be recognized by AP2. Thus, after activation, endocytosis will be inhibited and ectocytosis will be favored at the immune synapse.

Our findings point to two important roles for ectocytosis. First, ectocytosis removes activated TCRs from the membrane of CTLs, terminating signaling as ectosomes are taken up by dying targets. Second, TCR ectocytosis allows CTL and target to disengage, with areas

of separation arising naturally as ectosomes bud off from the CTL membrane. Although these areas might be small during ongoing TCR recognition of the target, as target recognition ceases, ectocytosis would allow complete separation and detachment, facilitating serial killing.

Here we show that activation-induced membrane specialization switches biological function within the immune synapse. These findings suggest that TCR signaling is self-limiting through a process that seamlessly links receptor signaling with shedding of activated TCRs and CTL detachment. We propose that this may provide a general mechanism for shedding receptors after activation that is likely to be used in many biological systems, including cilia (*28–32*).

REFERENCES AND NOTES

1. A. H. Courtney, W. L. Lo, A. Weiss, *Trends Biochem. Sci.* **43**, 108–123 (2018).
2. A. Alcover, B. Alarcón, V. Di Bartolo, *Annu. Rev. Immunol.* **36**, 103–125 (2018).
3. N. Blanchard *et al*., *J. Immunol.* **168**, 3235–3241 (2002).
4. K. Choudhuri *et al*., *Nature* **507**, 118–123 (2014).
5. D. G. Saliba *et al*., *eLife* **8**, e47528 (2019).
6. J. D. Martell *et al*., *Nat. Biotechnol.* **30**, 1143–1148 (2012).
7. M. R. Jenkins *et al*., *eLife* **3**, e01310 (2014).
8. C. R. Monks, B. A. Freiberg, H. Kupfer, N. Sciaky, A. Kupfer, *Nature* **395**, 82–86 (1998).
9. J. M. Stein, J. P. Luzio, *Biochem. J.* **274**, 381–386 (1991).
10. J. C. Stinchcombe *et al*., *J. Cell Biol.* **152**, 825–834 (2001).
11. M. R. Jenkins *et al*., *J. Exp. Med.* **212**, 307–317 (2015).
12. C. M. Gawden-Bone *et al*., *Immunity* **49**, 427–437.e4 (2018).
13. D. Allan, P. Thomas, R. H. Michell, *Nature* **276**, 289–290 (1978).
14. D. Allan, R. H. Michell, *Nature* **258**, 348–349 (1975).
15. S. Carrasco, I. Merida, *Mol. Biol. Cell* **15**, 2932–2942 (2004).
16. H. R. Melowic *et al*., *J. Biol. Chem.* **282**, 21467–21476 (2007).
17. A. T. Ritter *et al*., *Immunity* **42**, 864–876 (2015).
18. S. Kumari *et al*., *EMBO J.* **39**, e102783 (2020).
19. D. A. Cantrell, A. A. Davies, M. J. Crumpton, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8158–8162 (1985).
20. S. Valitutti, S. Müller, M. Cella, E. Padovan, A. Lanzavecchia, *Nature* **375**, 148–151 (1995).
21. S. Valitutti, S. Müller, M. Dessing, A. Lanzavecchia, *J. Exp. Med.* **183**, 1917–1921 (1996).
22. S. Valitutti, S. Müller, M. Dessing, A. Lanzavecchia, *Eur. J. Immunol.* **26**, 2012–2016 (1996).
23. S. Valitutti, S. Müller, M. Salio, A. Lanzavecchia, *J. Exp. Med.* **185**, 1859–1864 (1997).
24. B. T. Kelly *et al*., *Science* **345**, 459–463 (2014).
25. N. De Franceschi *et al*., *Nat. Struct. Mol. Biol.* **23**, 172–179 (2016).
26. L. M. Traub, J. S. Bonifacino, *Cold Spring Harb. Perspect. Biol.* **5**, a016790 (2013).
27. A. L. Szymczak, D. A. Vignali, *J. Immunol.* **174**, 4153–4160 (2005).
28. M. Cao *et al*., *eLife* **4**, e05242 (2015).
29. A. R. Nager *et al*., *Cell* **168**, 252–263.e14 (2017).
30. S. C. Phua *et al*., *Cell* **168**, 264–279.e15 (2017).
31. G. Garcia 3rd, D. R. Raleigh, J. F. Reiter, *Curr. Biol.* **28**, R421–R434 (2018).
32. M. V. Nachury, D. U. Mick, *Nat. Rev. Mol. Cell Biol.* **20**, 389–405 (2019).

ACKNOWLEDGMENTS

We thank J. P. Luzio, J. Kaufman, A. C. Richard, T. Douanne, and previous lab members for helpful discussions; S. Vaughan for tomography advice; E. Latimer-Bowman for serial sectioning; and the Cambridge Institute for Medical Research (CIMR) core imaging facility for assistance with super-resolution imaging. **Funding:** This work was funded by grants to G.M.G. from the Wellcome Trust (100140 and 217100). FIB-SEM imaging was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (CC0199), the UK Medical Research Council (CC0199), and the Wellcome Trust (CC0199). A.N. acknowledges funding by the European Research Council under the European Union's Horizon 2020 research and innovation program, grant agreement GA758334 ASYMMEM; the Max Planck Lipid Research Center; and the Deutsche Forschungsgemeinschaft (DFG) as a member of the TRR83 consortium. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. **Author contributions:** J.C.S. and G.M.G. conceived and wrote this study. J.C.S. carried out all experimental work with the following exceptions: Y.A. carried out live imaging of ectosome release; C.J.G.K. contributed to modeling; L.M.C. and C.J.P. performed FIB-SEM imaging; and K.B. and A.N. synthesized and gave advice for 1-stearoyl-2-arachidonoyl-sn-glycerol (cgSAG) and DAG experiments. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.apb8933
Materials and Methods
Figs. S1 to S11
Movies S1 to S9
References (*33–36*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 4 March 2022; resubmitted 10 February 2023
Accepted 28 April 2023
10.1126/science.abp8933

SOLAR CELLS

Lead-chelating hole-transport layers for efficient and stable perovskite minimodules

Chengbin Fei[1], Nengxu Li[1], Mengru Wang[1], Xiaoming Wang[2], Hangyu Gu[1], Bo Chen[3], Zhao Zhang[3], Zhenyi Ni[1], Haoyang Jiao[1], Wenzhan Xu[1], Zhifang Shi[1], Yanfa Yan[2], Jinsong Huang[1,4]*

The defective bottom interfaces of perovskites and hole-transport layers (HTLs) limit the performance of p-i-n structure perovskite solar cells. We report that the addition of lead chelation molecules into HTLs can strongly interact with lead(II) ion (Pb^{2+}), resulting in a reduced amorphous region in perovskites near HTLs and a passivated perovskite bottom surface. The minimodule with an aperture area of 26.9 square centimeters has a power conversion efficiency (PCE) of 21.8% (stabilized at 21.1%) that is certified by the National Renewable Energy Laboratory (NREL), which corresponds to a minimal small-cell efficiency of 24.6% (stabilized 24.1%) throughout the module area. Small-area cells and large-area minimodules with lead chelation molecules in HTLs had a light soaking stability of 3010 and 2130 hours, respectively, at an efficiency loss of 10% from the initial value under 1-sun illumination and open-circuit voltage conditions.

The power conversion efficiency (PCE) of small-area n-i-p structure, single-junction perovskite solar cells has reached the parity of monocrystalline silicon solar cells (>25%) (*1*–*4*). However, the certified perovskite module PCE is still much lower, at about 19% (*5*–*7*), because of the low PCE of p-i-n structure perovskite solar cells and large losses after transferring the small cells to modules, which is mainly induced by the nonuniformity of perovskites or charge-transport layers. Several strategies, such as improving the crystallization process and passivating surface defects, have been applied to further improve the efficiency of p-i-n structure devices (*8*–*12*). For example, a certified stabilized PCE >24% was reported for small-area p-i-n structure devices when a surface treatment was used to enhance charge extraction (*13*). In contrast to top surfaces, the perovskite–hole-transport layer (HTL) interface has received much less attention. Nonradiative recombination of interface carriers caused by defects at buried interfaces may limit the efficiency of p-i-n structure perovskite solar cells (*14*). A notable smaller photoluminescence (PL) quantum yield was observed for perovskites near HTLs (*15*).

In contrast to top surfaces, the bottom perovskite-HTL interface is much more difficult to control. Attempts have been made to mix defect passivation species with HTLs (*16*–*19*), given that many passivation molecules may be washed away during perovskite coating because of their high solubility in perovskite precursor solvents, such as 2-methoxyethanol (2-ME), dimethylformamide (DMF), and dimethyl sulfoxide (DMSO). The other challenge comes from the top-to-down crystallization process of perovskites during solution coating (*20*). The evaporation of initially trapped DMSO may generate amorphous regions and voids on the scale of tens of nanometers at the bottom of perovskite films during film formation (*6*). Although the initially formed voids could be eliminated by introducing solid-state carbohydrazide to partially replace DMSO in perovskite precursor solutions, new voids formed at the bottom interfaces along grain boundaries during long-term device operation, indicating the presence of non-negligible amorphous perovskite in perovskite films even after annealing (*21*). Amorphous regions near the bottom interface have been also observed by high-resolution tunneling electron microscopy studies (*22*, *23*). The nonuniform distribution of the amorphous regions, caused by evaporation of DMSO, may cause nonuniform device performance, and thus it strongly affects the module efficiency and reproducibility, which is not reflected by champion small-cell PCEs. Perovskite modules need to be fabricated to evaluate their impact.

We report a method to effectively reduce the amorphous region at the bottom of perovskite films by embedding lead chelation molecules (LCMs), including a broadly applied electron transport material of bathocuproine (BCP), into HTLs. BCP competed with DMSO to interact with lead ions through strong chelation, which reduces the DMSO residue and thus the amorphous region in perovskites near the HTLs. BCP also passivated perovskites and improved the PCE, reproducibility, and stability of the perovskite cells and minimodules.

Strong interaction of LCMs with perovskites

We passivated the bottom perovskite-HTL interface by blending passivation species into HTLs. We used LCMs, which are used in water purification or as drugs to treat lead poisoning (*24*), because these molecules should interact strongly with Pb^{2+} in perovskites at their interface with HTLs (Fig. 1A). The molecular structures of several selected LCMs are shown in figs. S1 and S2, including BCP, *p*-toluenesulfonic anhydride (TSA), L-ascorbic acid (VC), meso-2,3-dimercaptosuccinic acid (MDSA), and ethylenediaminetetraacetic acid (EDTA), all of which can provide two or more lone-pair electrons to chelate with Pb^{2+} on perovskite surfaces. The chelation of LCMs with Pb^{2+} led to the precipitation of Pb^{2+} after addition to the perovskite precursor solution. We selected BCP for demonstration because of its low solubility of <0.2 mg ml^{-1} in 2-ME, the main solvent in our perovskite inks for a fast-blading process. When BCP was mixed with formamidinium lead iodide ($FAPbI_3$) solution, pale yellow precipitates showed up even at a very low BCP concentration of <0.05 mg ml^{-1} (Fig. 1B), indicating that the reaction products have an even lower solubility than BCP. The same precipitation phenomenon was observed when TSA was blended with $FAPbI_3$ (fig. S3).

The x-ray diffraction (XRD) pattern of the pale yellow precipitates (Fig. 1C and fig. S4) revealed diffraction peaks of multiple compounds, including PbI_2•BCP and some solvated phases. The diffraction peak at 11.06° was assigned to PbI_2•BCP because this peak was observed from the precipitates of the PbI_2 and BCP mixture solution (fig. S5). However, all of the solvated phases, such as PbI_2•DMSO, disappeared after removing DMSO by annealing the precipitates at 100°C, after which the powder color changed from yellow to pale yellow. The presence of PbI_2•BCP in the annealed powder indicated that the interaction of BCP with Pb^{2+} was stronger than that of DMSO with Pb^{2+}.

Fourier-transform infrared spectroscopy (FTIR) measurements were conducted to study how BCP interacts with perovskite precursors (Fig. 1, D and E). The C=N stretch in BCP at 1622 cm^{-1} could be distinguished readily from that of $FAPbI_3$ at 1720 cm^{-1} (*25*). After mixing with $FAPbI_3$, the C=N stretching peak from BCP shifted to a smaller wavenumber of ~1606 cm^{-1}, which agreed with the shifting of electron density toward Pb^{2+} and was indicative of the strong interaction of BCP with Pb^{2+}. The same FTIR shift was observed when all of the other LCMs were mixed with $FAPbI_3$ (fig. S6), supporting the Pb^{2+} chelation effect of these molecules. The phenanthroline group in BCP has two pyridine rings. We note that the C=N FTIR peak completely shifted, rather than splitting into two peaks. This result indicated that the N atom in both pyridine groups of BCP interacted with Pb^{2+}, resulting in the chelation of BCP with Pb^{2+}. X-ray photoelectron spectroscopy (XPS) also confirmed the interaction of BCP with Pb^{2+}, because there was a shift in the Pb 4f peak after the

[1]Department of Applied Physical Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [2]Department of Physics and Astronomy, University of Toledo, Toledo, OH 43606, USA. [3]Perotech, Inc., Chapel Hill, NC 27516, USA. [4]Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
*Corresponding author. Email: jhuang@unc.edu

perovskite film was covered by a spin-coated thin layer of BCP (Fig. 1F). The decrease in Pb^{2+} binding energy was consistent with the chelation of BCP with Pb^{2+} that resulted from the sharing of electrons from N (in BCP) with Pb^{2+}.

The chelation effect provides a more robust interaction with perovskites than the single pyridine group. We calculated the binding energies of a BCP and a DMSO molecule on PbI_2-terminated and formamidinium iodide (FAI)–terminated (100) $FAPbI_3$ perovskite surfaces. During the PbI_2 termination, the two N atoms of BCP bond to a single Pb atom of $FAPbI_3$ (Fig. 1G), whereas for DMSO, only a single O–Pb bond forms (Fig. 1H). The binding energy of BCP with $FAPbI_3$ is 1.71 eV and 0.84 eV for DMSO. For the FAI termination, hydrogen bonding occurs between the two N atoms of BCP and one H atom of FA (Fig. 1I). The binding energy of BCP with $FAPbI_3$ is 1.41 eV. For DMSO, hydrogen bonding occurs between the O atom of DMSO and two H atoms of FA^+ (Fig. 1J), resulting in a smaller binding energy of 0.57 eV. On both surfaces with PbI_2 and FAI termination, the binding energy is two to three times greater for BCP than for DMSO.

Defect passivation of the bottom interface by LCMs

We introduced LCMs to the poly[bis(4-phenyl)(2,4,6-trimethylphenyl)amine] (PTAA) layer by mixing them in toluene at different ratios for BCP and TSA or by blade-coating them onto the PTAA layers for insoluble LCMs, including VC, MDSA, and EDTA (see supplementary methods for details). The perovskite composition of $FA_{0.9}Cs_{0.1}PbI_3$ was used for device fabrication because of its reported high operational stability (*5*, *26*–*28*). Both the HTL and perovskite layers were coated in ambient conditions by using nitrogen knife-assisted blading (*29*).

We conducted steady-state PL and PL lifetime measurements to understand how LCMs blended in PTAA affect the perovskite-HTL interfaces. Several prior studies showed that the embedded perovskite-HTL interface could be even more defective than the top perovskite surface depending on the processing conditions (*15*). When we sandwiched a perovskite film with PTAA to keep the charge extraction condition the same at both sides and compared the PL intensity of the top surface with that of the bottom surface, we also observed a ~ 40% weaker PL from the bottom side (fig. S7), indicating that bottom interface indeed has more defects. When we examined the bottom interface by bringing incident excitation light in from the indium tin oxide (ITO) side (Fig. 2A), BCP notably enhanced the steady-state PL intensity by 1.7 times (Fig. 2B). The PL lifetime at the bottom of the perovskite films also increased from 195 to 452 ns (Fig. 2C), so preembedded BCP in PTAA reduced the nonradiative recombination defects at the bottom interface of perovskites. However, the improved PL lifetime may also be caused by the HTL creating a more-crystalline perovskite layer (*30*). To verify the passivation function of BCP, we spun a BCP layer on top of perovskites to exclude the impact of perovskite crystallinity. We measured the PL spectra and PL lifetime changes with light incident from the air side (Fig. 2D). BCP enhanced the PL intensity by 1.2 times and improved the PL lifetime to 1.43 μs (Fig. 2, E and F), which confirmed the passivation effect of BCP on perovskites and indicated that the

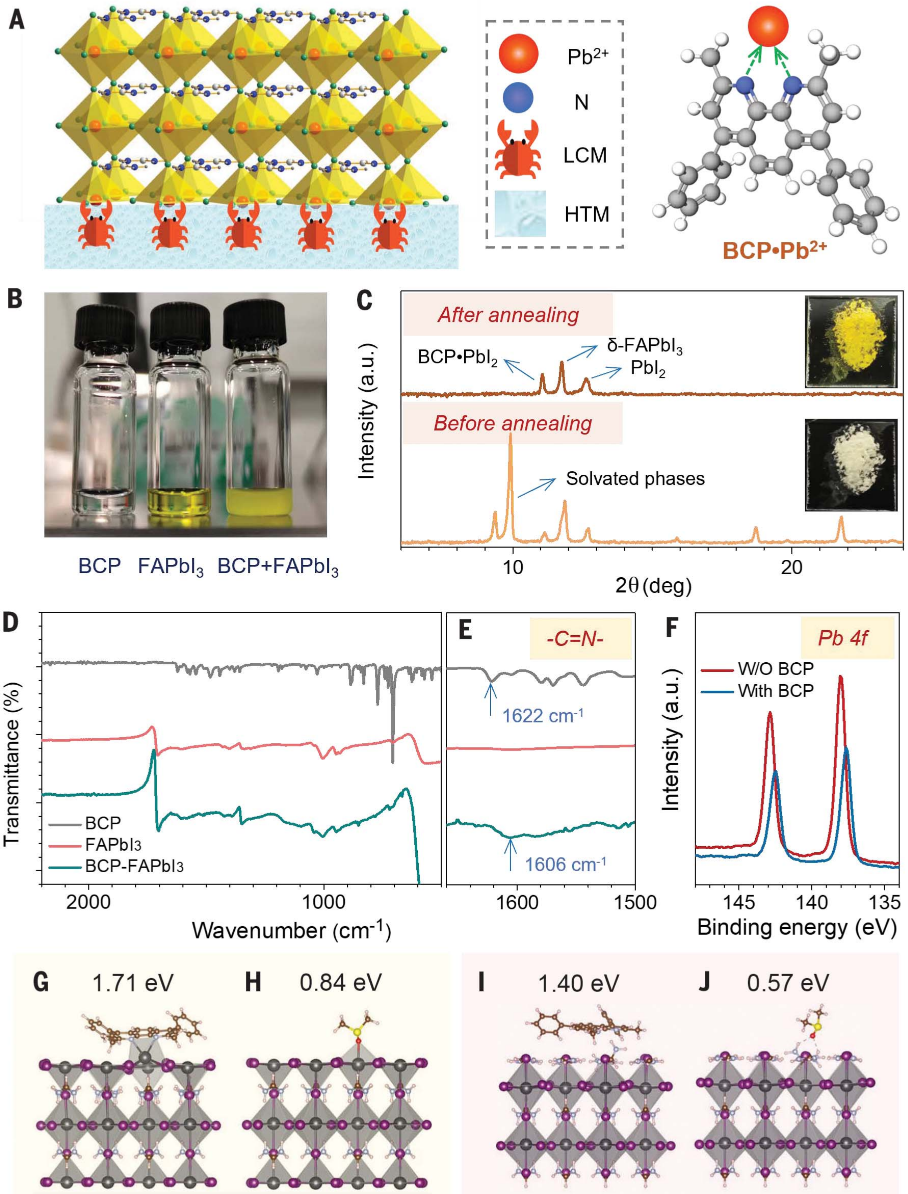


Fig. 1. Chelation of lead ions by BCP. (**A**) Illustration of the chelation of the LCMs (in the HTL) with Pb^{2+} ions at the bottom side of the perovskite films. (**B**) Photo of the BCP solution (in 2-ME, ~0.1 mg ml^{-1}), $FAPbI_3$ solution (in 2-ME and DMSO, 1M), and a mixture of the BCP solution and the $FAPbI_3$ solution (volume: volume = 1:1 mixture). (**C**) XRD pattern of the precipitates that form in the mixture of the BCP and $FAPbI_3$ solutions. The insets are photos of the dried precipitates before (bottom) and after (top) annealing. a.u., arbitrary units. (**D** and **E**) FTIR of the BCP powder, the $FAPbI_3$ powder, and the dried powder from the BCP-$FAPbI_3$ mixture (D). A zoomed-in view of the FTIR results is shown in (E). (**F**) XPS spectra of Pb 4f from the perovskite films with and without spun BCP on the top surface. (**G** to **J**) DFT modeling of the interaction of the BCP and PbI_2 termination (G), the DMSO and PbI_2 termination (H), the BCP and FAI termination (I), and the DMSO and FAI termination (J). [Photo credits: The authors]

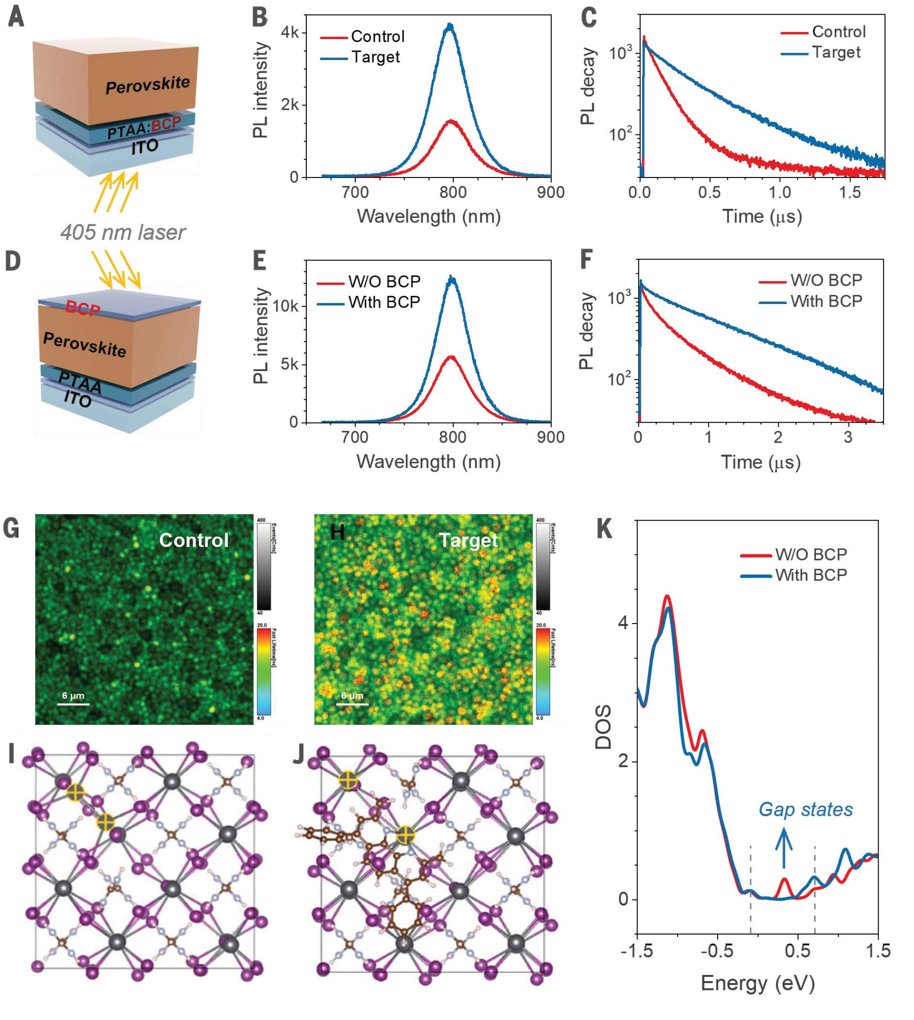


Fig. 2. Functionality of BCP in reducing defects in perovskites. (**A** and **D**) Measurement geometry. (**B** and **C**) Steady state PL and time-resolved PL (TRPL) decay of control and target perovskite films, respectively. (**E** and **F**) Steady-state PL and TRPL decay of perovskite samples with and without BCP coated on the top surface, respectively. (**G** and **H**) FLIM of perovskite layers on PTAA with and without BCP, respectively. (**I**) Structure of a Pb-Pb dimer on the PbI_2-terminated (100) surface. (**J**) Structure of BCP binding to a Pb-Pb dimer on the surface. (**K**) Calculated DOS of the PbI_2 termination without and with BCP binding. The dashed vertical lines indicate the valence band maximum and conduction band minimum.

BCP addition also modified the morphology of the perovskite close to this interface.

We evaluated the uniformity of the passivated perovskite-HTL interface using fluorescence lifetime imaging microscopy (FLIM). Hereafter, we refer to the pure PTAA and BCP: PTAA HTLs as control and target, respectively. The crystal grains were distinguished in PL mapping and had apparent sizes of 1 to 2 μm (Fig. 2, G and H). The PL intensity and lifetime were relatively uniform in both the control and target films over a measurement area of 40 μm by 40 μm, but compared with the control film, the target film showed a PL that was 1.5 times stronger and a PL lifetime that was 1.1 times longer.

We further compared the passivation effects of BCP and DMSO by density function theory (DFT) calculation. We found that perfect PbI_2- and FAI-terminated surfaces do not form deep gap states. Pb-Pb dimers could form on the surface with an extra I vacancy, which can be easily formed through the loss of iodide. The Pb-Pb dimer induces gap states through Pb 5p–Pb 5p bonding. The deep gap states could be eliminated by BCP but not by DMSO, as revealed by studies of the PbI_2-terminated (100) surface. The Pb-Pb dimer that formed when two iodide vacancies were present created deep gap states (Fig. 2I). When a BCP molecule interacted with the Pb-Pb dimer, the two N atoms bonded to one Pb atom of the dimer, breaking the dimer structure (Fig. 2J), because the Pb-N bond formation partly saturated the Pb coordination. As shown in Fig. 2K, the gap states were eliminated after breaking the Pb-Pb dimer. However, DMSO could not break Pb-Pb dimers because it formed only one Pb–O bond. The Pb-Pb dimer remained after DMSO molecule binding, leaving gap states (fig. S8, A and B).

Reducing amorphous regions in perovskites by LCMs

To find out which defects were reduced by the introduction of BCP to PTAA, we measured the trap density change of the devices using thermal admittance spectroscopy and drive-level capacitance profiling (DLCP). The trap density of state (tDOS) profiles in Fig. 3A showed three distinct trap bands for both the control and target devices, which resembled those reported for iodide perovskites. These trap bands were assigned to negatively charged iodine interstitials (I_i^-) and positively charged iodine interstitials (I_i^+), respectively (*31*). The density of I_i^- did not exhibit obvious changes in the perovskites, whereas the I_i^+ density was reduced by the introduction of BCP into PTAA. DLCP measurements gave consistent results in that the I_i^- density did not change throughout the devices, whereas the I_i^+ density was notably reduced (fig. S9).

We plotted the change in I_i^+ density ($\Delta I_i^+ = |I^+_{i,\text{target}} - I^+_{i,\text{control}}|$) induced by BCP in perovskites across the device from C_{60} to the HTL side. As shown in Fig. 3B, BCP mainly reduced the I_i^+ density in perovskites near the HTL. Our recent study of another perovskite with the same composition showed a similar reduction in the I_i^+ density with unchanged I_i^- density upon thermal annealing of the films in the dark for several hundreds of hours (*21*). Because the reduction in the I_i^+ density was accompanied by a reduction of the amorphous region and the appearance of associated voids at the bottom interface near PTAA, we hypothesize that the amorphous phase at the bottom of the perovskite films may be rich in I_i^+, which were reduced by the BCP:PTAA HTL.

To verify this hypothesis, we conducted scanning electron microscope (SEM) measurements to examine the voids caused by the recrystallization of the amorphous region in the control perovskite devices using PTAA as the HTL. The devices were encapsulated to avoid the impact of moisture and oxygen and soaked under room lighting for ~1000 hours to allow recrystallization to occur. The fresh and soaked perovskite films were peeled off from the ITO/PTAA substrate to expose the bottom surfaces, which were then subjected to SEM study. As shown in Fig. 3, C to F, the average grain size of these perovskite films was 1 to 2 μm, which agreed with the PL mapping result. For the fresh samples with or without BCP, there was no notable morphology difference at the bottom of the perovskite films. After aging, many shallow voids appeared

around the grain boundaries at the bottom of the aged control samples, resembling the result of thermal annealing of such films (*21*).

By contrast, the bottom morphology of the aged target sample was nearly indistinguishable from that of the fresh target samples, indicating that BCP notably reduced the formation of an amorphous phase at the bottom of the perovskite films. We conducted grazing-incident XRD (GIXRD) pattern measurements to directly examine the crystallinity of perovskites near the HTL using peeled-off fresh samples. As shown by the GIXRD patterns in Fig. 3G and fig. S10, the peak of the (100) plane is not only 60% stronger but also 8% sharper, confirming the improved crystallinity of the perovskite bottom layer induced by BCP. The reduced amorphous regions should contribute to the reduction in I_i^+ density at the bottom of the perovskite films (Fig. 3A).

We studied why BCP could reduce the amorphous regions. Prior studies showed that the trapped solvent DMSO may also cause the amorphous phase in the perovskite bottom and grain boundaries (*6*). We measured the trapped DMSO content in perovskite films after film annealing using hydrogen nuclear magnetic resonance (^{1}H-NMR) spectroscopy. Perovskite films fabricated on different HTLs were scraped off from the substrates and dissolved in deuterium oxide. The content of DMSO in the perovskite film was determined by the ratio of the integral area of the characteristic peak of ^{1}H in DMSO to that of FA^+. As shown in Fig. 3H and fig. S11, the peaks at 7.71 parts per million (ppm) and 2.63 ppm were assigned to –CH in FA^+ and $–CH_3$ in DMSO, respectively (*32*, *33*). The DMSO/FA^+ atomic ratio in the control sample was 0.13% but was 0.08% in the target sample (table S1).

Thus, BCP in PTAA did reduce the amount of trapped DMSO, in that BCP competed with DMSO to coordinate with Pb^{2+} at the bottom region during the film formation process, so BCP:PTAA reduced the amount of DMSO bounded to Pb^{2+} (Fig. 1C). The decrease in trapped DMSO resulted in perovskites that were less amorphous. The reduction of defective amorphous perovskite by BCP thus improved device efficiency and stability. The trapping of DMSO can be very nonuniform, particularly for large-area perovskite films coated in a production line, depending on the local vapor pressure, which could differ substantially from location to location. Reduced trapping of DMSO explained the better uniformity of perovskite films and thus would be expected to enhance perovskite module efficiency.

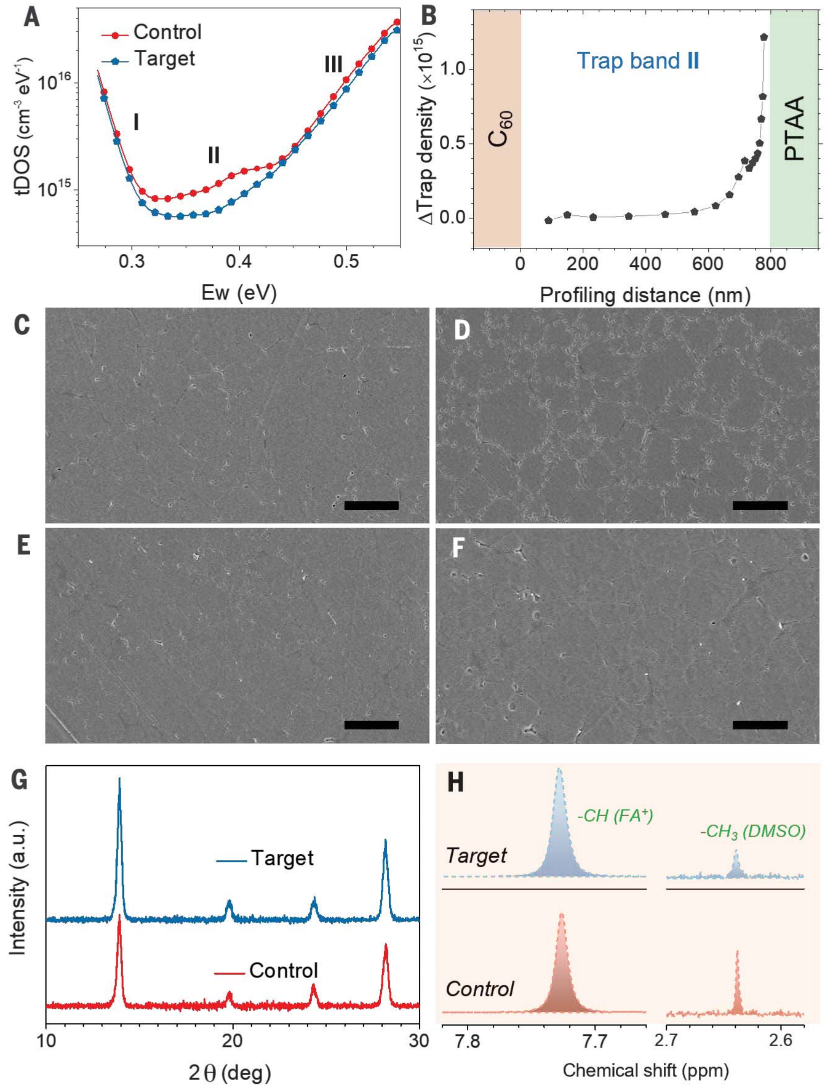


Fig. 3. Influence of BCP on perovskite crystallization. (**A**) tDOS of perovskite solar cells prepared on PTAA with (target) and without (control) BCP. I, II, and III represent trap bands. E_w, demarcation energy. (**B**) Changes in the trap density of trap band II in the perovskite devices induced by BCP. SEM images of the bottom side of the perovskite films that were peeled off from the ITO/glass substrates. (**C** to **F**) SEM images of the control films [(C) and (D)] and target films [(E) and (F)]. (C) and (E) are fresh films, and (D) and (F) are films aged 1000 hours. Scale bars are 2 μm. (**G**) GIXRD patterns of the fresh control and target perovskite films measured from the bottom side. (**H**) ^{1}H-NMR of DMSO residual in the control and target perovskite films.

Enhancing the performance of perovskite solar cells by LCMs

To evaluate how the LCMs affect perovskite solar cell performance, perovskite solar cells were fabricated with the structure ITO/HTL/perovskite/C_{60}/BCP/copper (Cu) (Fig. 4A). As shown in fig. S12, incorporating LCMs either into or onto HTLs improved the stability of all of the devices during an accelerated light-induced degradation test (200 mW cm^{-2}, at open-circuit and ambient conditions, 59° to 65°C). Of all of the LCMs tested, BCP and TSA resulted in a higher PCE with an improved open-circuit voltage (V_{OC}) and fill factor (FF) (fig. S13). The higher PCEs could be explained by the better mixing of BCP and TSA with PTAA, because the benzene group in BCP or TSA should induce a π-π interaction with PTAA (*34*), facilitating carrier transfer from perovskites to PTAA. Another reason is that BCP and TSA are much less soluble in the solvents of perovskites.

We chose BCP for the PCE optimization. BCP has been broadly used as an electron transport material in p-i-n structure perovskite solar cells (*35*), and its hole-blocking property has been reported often (*36*–*38*). Thus, its loading ratio in PTAA needed to be optimized for optimal hole collection. Perovskites have a sufficiently long carrier diffusion length, such that photogenerated charges can still be efficiently extracted even if some local contacts are blocked, which has been demonstrated by nanocontact perovskite solar cells or those

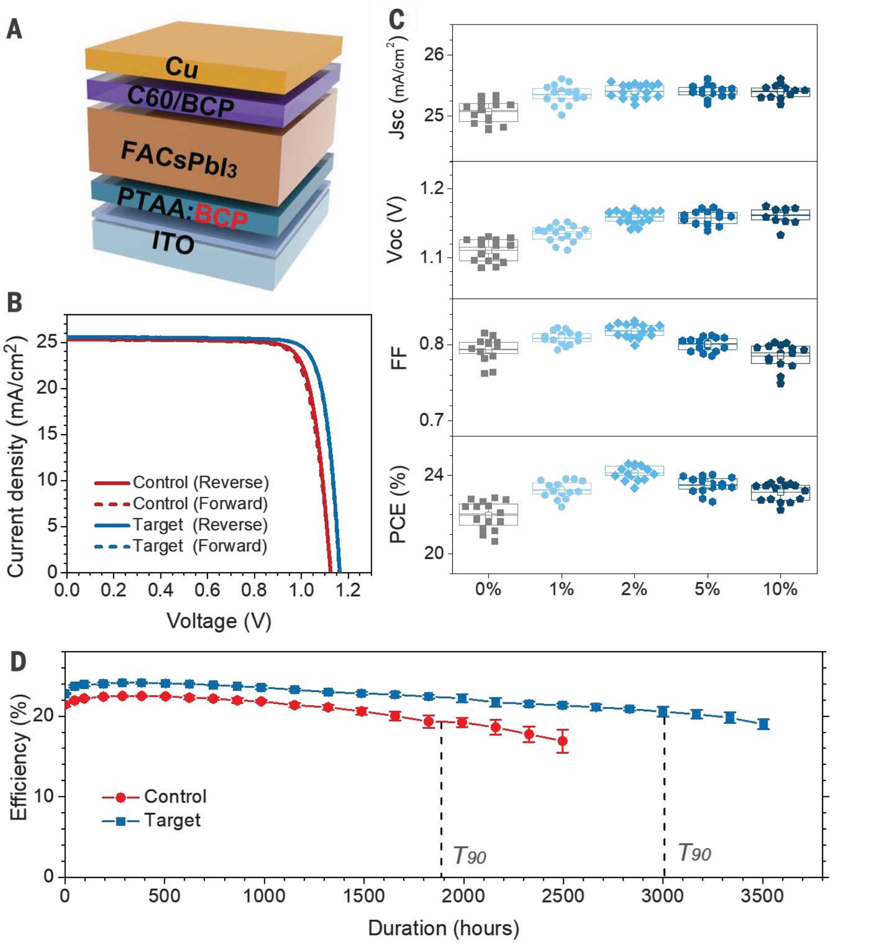


Fig. 4. Small-area perovskite device performance. (**A**) Schematic of the structure of the perovskite solar cells. (**B**) *J-V* curves of the p-i-n structure small devices with and without BCP in PTAA HTLs. The active area of the small devices is 0.08 cm^2. (**C**) Device performance parameter values for the small devices based on PTAA with 0 to 10 wt % BCP in PTAA. Each of the symbols (squares, circles, diamonds, hexagons, and pentagons) represent one sample. The box outlines represent the data variation. (**D**) Light soaking stability of the control and target small devices under illumination from light source LS2 with an intensity of 100 mW cm^{-2}. Data were collected from 18 to 20 devices for each group.

with noncontinuous insulating interfacial layers (*19*, *39*–*42*). The control devices with a 0.08-cm^2 working area had a short-circuit current density (J_{SC}) of 25.2 mA cm^{-2}, a V_{OC} of 1.13 V, a FF of 0.80, and a PCE of 22.7% (Fig. 4B), consistent with our previous results (*5*). As shown in Fig. 4C, a BCP of ~2 wt % in PTAA was needed to increase the FF of the perovskite devices. When the BCP loading in PTAA exceeded 2 wt %, the device FF was markedly reduced, which was likely caused by a reduction in hole collection. The same variation trend of J_{SC} with that of FF also occurred with varied BCP loading ratios. By contrast, the device V_{OC} monotonically increased from 1.13 to 1.17 V when the BCP loading ratio was increased from 0 to 2 wt %. V_{OC} saturated with a further increased BCP ratio because hole extraction was not needed under open-circuit conditions.

The champion target devices with optimal BCP showed a J_{SC} of 25.5 mA cm^{-2}, a V_{OC} of 1.17 V, and a FF of 0.825, resulting in a PCE of 24.6% (Fig. 4B). The average PCE increased to 24.1%, and the photocurrent hysteresis was not obvious in either type of device. The observed small variation in the PCE of small-area devices (0.08 cm^2) agreed with the uniform distribution of BCP in PTAA and uniform passivation. The average PCE of the control devices is ~22.0%, which is consistent with what was reported previously using the same CsI- or FAI-rich perovskite composition (*5*). Considering that BCP has other functional groups besides phenanthroline, we evaluated two other molecules with these functional groups and confirmed that chelation phenanthroline groups dominate the device efficiency enhancement (figs. S14 and S15).

The preembedded BCP in PTAA also greatly improved the device stability. We used three different light sources for stability measurements based on the device area and lifetime of the light sources (fig. S16). The light intensity of all of the light sources was controlled to be 1 sun, as determined by a silicon solar cell that is used for solar-simulator light-intensity calibration (see fig. S17). The xenon light source with a 400-nm long-pass filter (LS1) and the white light-emitting diode (LED) light source (LS2) had more deep-blue or deep–ultraviolet (UV) light, and the light-emitting plasma light source with a 450-nm long-pass filter (LS3) had a stronger infrared component. Deep-UV light was filtered to simulate the final PV modules that have a UV absorber such as a ZnO coating (*43*, *44*). We also evaluated whether the spectra of these light sources affected solar-cell stability. Our study shows that UV light in the solar spectrum did accelerate perovskite device degradation, but filtering UV light of the solar spectrum made perovskites as stable as those measured under a white LED light source (fig. S18). The degradation of these devices was similar (fig. S19), and when the deep-UV light was filtered, only minor spectral differences in the stability of the perovskite solar cells were evident for the three light sources. The small-area device stability was tested by soaking the devices at open-circuit conditions with light source LS2. The devices were heated to between 55° and 61°C by illumination during the stability testing, and no additional temperature control was applied.

Typical *J-V* curves of the devices after light soaking for different durations are shown in fig. S20, and average PCEs over time for the control and target devices are shown in Fig. 4D. For the control devices, the PCE decreased to 90% of its initial efficiency (T_{90} lifetime) after light soaking for 1890 hours. By contrast, the T_{90} lifetime of the target devices under light soaking increased to 3010 hours. Previous studies concluded that light-soaking stability measured at V_{OC} conditions is an even harsher stability test than the maximum power point (MPP) stability test because accumulated photogenerated excess charges accelerate perovskite degradation by enhancing ion migration (*45*). To verify this for this perovskite composition, we tested the same batch of control devices under the MPP and V_{OC} conditions and found a much longer T_{90} lifetime during MPP testing (fig. S21). The thermal stability of the devices was also evaluated at high temperatures according to the International Electrotechnical Commission (IEC) 61215 standard. The PCE loss of both the control and target devices was <1% of the initial efficiency (T_{99}) after ~1000 hours of testing in the dark at 85°C (fig. S22).

Efficiency and stability of perovskite minimodules

Our prior study showed that nonuniform photocurrents from different subcells would

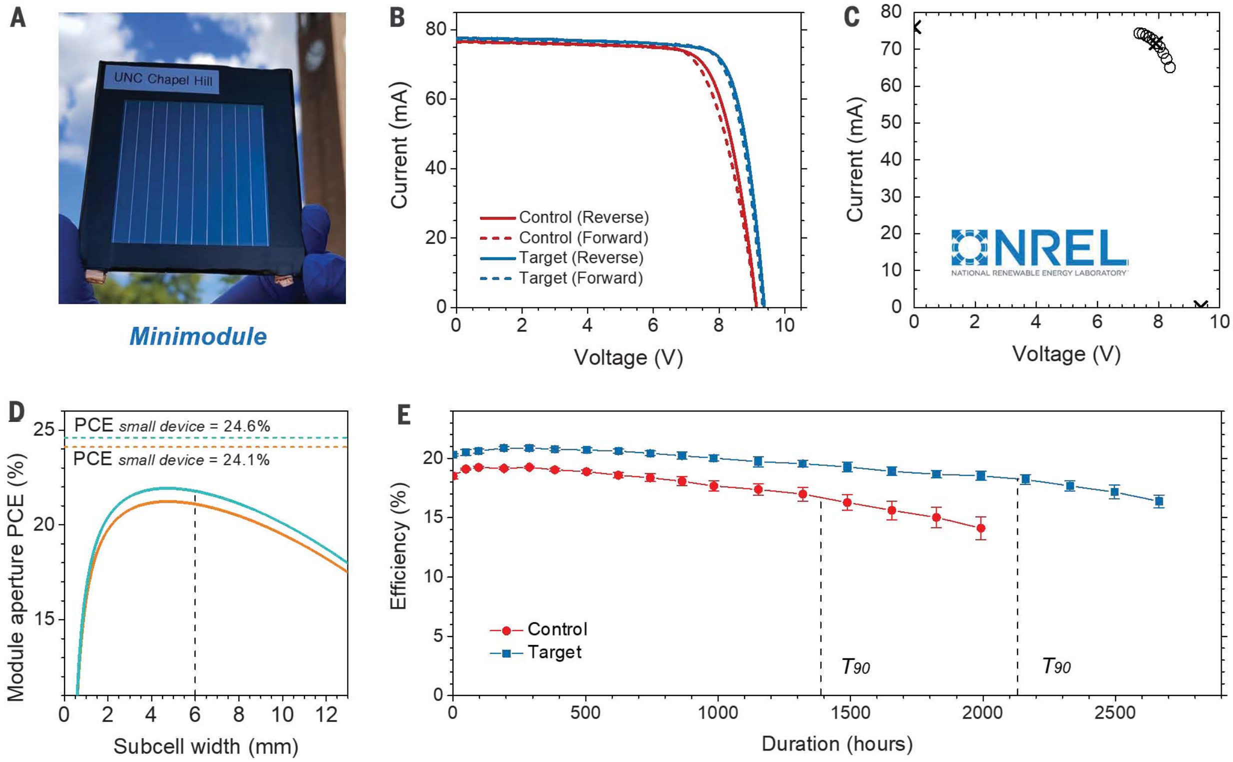


Fig. 5. Perovskite minimodule performance. (**A**) Photo of an encapsulated minimodule. [Photo credit: The authors] (**B**) *J-V* curves of the control and target minimodules that have eight subcells connected in series. Each subcell is 6 mm wide and 55 mm long. (**C**) Certification of stabilized photovoltaic performance for a minimodule by NREL using their asymptotic Pmax scan protocol. The aperture area of this module is 26.9 cm^2. (**D**) The calculated aperture efficiency of minimodules with different subcell widths based on small-cell efficiencies of 24.1% (orange) and 24.6% (green). The maximum attainable efficiencies for a subcell width of 6 mm are 21.1 and 21.8%. (**E**) Light-soaking stability of the control and target minimodules under illumination from light source LS2 with an intensity of 100 mW cm^{-2}. Efficiency data were collected from five or six minimodules (aperture area of ~26.4 cm^2) for each group.

notably reduce the FF of resultant modules (*46*). BCP:PTAA improved the uniformity of perovskite films, as evidenced by the small variation in device performance from batch to batch, which should also improve the module performance. We thus upscaled the 0.08 cm^2 small-area devices to minimodules with aperture areas of 20 to 30 cm^2, as defined by the photomask shown in fig. S23. The minimodules had a subcell width of 6.0 mm and a dead-area width of 320 μm (fig. S24), resulting in a geometry filling factor of 94.7% (Fig. 5A). As shown by the photocurrent curves in Fig. 5B, the addition of BCP enhanced the FF and V_{OC} of the minimodules. Several minimodules were sent to the National Renewable Energy Laboratory (NREL) for PCE certification. The certified champion aperture PCE of the perovskite minimodules reached 21.8%, with an aperture area of 26.9 cm^2 from *J-V* scanning (fig. S25). The stable power output measurement method (which measures the stabilized photocurrent near the MPP) gave a stabilized PCE of 21.1% (Fig. 5C and fig. S26).

Using a recently established model (*46*), we calculated the minimal small-cell PCE that is needed to yield a module PCE of 21.8% (or a stabilized PCE of 21.1%) by assuming that the perovskite device is perfectly uniform throughout the minimodules. As shown in Fig. 5D, with more details in fig. S27, the minimal small-area cell PCE needs to be 24.6% (stabilized PCE of 24.1%). This efficiency is near the measured average PCE of small cells, indicating that the perovskite films were very uniform. The minimodules with a BCP:PTAA HTL had a V_{OC} of 1.17 V, which is the same as that of small-area cells. The photocurrent density from each 6-mm-wide subcell in the minimodule was 23.6 mA cm^{-2}, which is near the calculated value (fig. S27A). Notably, the minimodule with this large area had a large FF of 0.803, which was also the highest FF reported for perovskite minimodules. The calculation result provided in fig. S27B also shows that the FF of 0.803 has already reached the theoretical upper limit based on the attainable small-area device performance, indicating that there is negligible PCE loss caused by the nonuniformity of device performance.

The minimodule stability was evaluated by light soaking with five or six minimodules in ambient conditions at V_{OC} conditions under light source LS2 with a light intensity of 1 sun. The minimodule temperature was measured as between 56° and 62°C and was kept constant during the testing process. Instead of using a higher testing temperature, the actual module operation temperature was used to keep the underlying photochemistry the same as it would be under real operation conditions. Typical *J-V* curves of the control and target minimodules after light soaking for different durations are shown in fig. S28. The average T_{90} lifetime of the minimodules with a BCP:PTAA HTL reached ~2130 hours, whereas the control minimodules only had a T_{90} lifetime of ~1380 hours (Fig. 5E).

Discussion

The introduction of LCMs into PTAA HTLs improved the efficiency, stability, and reproducibility of the p-i-n structure perovskite solar cells. Because of the strong interaction between the phenanthroline group in BCP and Pb^{2+} and the low solubility of BCP•PbI_2 in 2-ME, BCP stayed at the bottom of the perovskite layer during solution coating. With a reduced trap density at the bottom of the perovskite layer and charge recombination at the PTAA-perovskite interface, the average

PCE of the devices with BCP increased to 24.1%. With BCP in the PTAA layer, the amorphous layer and DMSO residue in the perovskite layer were decreased, leading to the reduced void formation around the grain boundaries during light soaking. BCP in PTAA also increased the T_{90} lifetime from ~1890 to ~3010 hours. The champion minimodules with an aperture area of 26.9 cm^2 showed a certified PCE of 21.8% (stabilized at 21.1%).

REFERENCES AND NOTES

1. M. Kim *et al*., *Science* **375**, 302–306 (2022).
2. Y. Zhao *et al*., *Science* **377**, 531–534 (2022).
3. T. Wang *et al*., *Science* **377**, 1227–1232 (2022).
4. H. Min *et al*., *Nature* **598**, 444–450 (2021).
5. Y. Deng *et al*., *Nat. Energy* **6**, 633–641 (2021).
6. S. Chen *et al*., *Science* **373**, 902–907 (2021).
7. T. Bu *et al*., *Science* **372**, 1327–1332 (2021).
8. Q. Cao *et al*., *Sci. Adv.* **7**, eabg0633 (2021).
9. Z. Li *et al*., *Science* **376**, 416–420 (2022).
10. X. Li *et al*., *Science* **375**, 434–437 (2022).
11. S. Bai *et al*., *Nature* **571**, 245–250 (2019).
12. S. Tan *et al*., *Nature* **605**, 268–273 (2022).
13. Q. Jiang *et al*., *Nature* **611**, 278–283 (2022).
14. X. Zuo *et al*., *Chem. Eng. J.* **431**, 133209 (2022).
15. G. Yang *et al*., *Nat. Photonics* **16**, 588–594 (2022).
16. Y. Yang *et al*., *Org. Electron.* **86**, 105873 (2020).
17. B. L. Watson, N. Rolston, K. A. Bush, L. Taleghani, R. H. Dauskardt, *J. Mater. Chem. A* **5**, 19267–19279 (2017).
18. S. Ye *et al*., *J. Am. Chem. Soc.* **139**, 7504–7512 (2017).
19. J. Peng *et al*., *Energy Environ. Sci.* **10**, 1792–1800 (2017).
20. S. Chen *et al*., *Sci. Adv.* **7**, eabb2412 (2021).
21. M. Wang, C. Fei, M. A. Uddin, J. Huang, *Sci. Adv.* **8**, eabo5977 (2022).
22. S. Yang *et al*., *Science* **365**, 473–478 (2019).
23. Z. Ni *et al*., *Science* **367**, 1352–1358 (2020).
24. J. J. Chisolm Jr., *J. Pediatr.* **73**, 1–38 (1968).
25. V. C. A. Taylor *et al*., *J. Phys. Chem. Lett.* **9**, 895–901 (2018).
26. J. Chen, N.-G. Park, *Adv. Mater.* **31**, e1803019 (2019).
27. N. Li *et al*., *Joule* **4**, 1743–1758 (2020).
28. Y.-H. Lin *et al*., *Science* **369**, 96–102 (2020).
29. Y. Deng *et al*., *Sci. Adv.* **5**, eaax7537 (2019).
30. C. Bi *et al*., *Nat. Commun.* **6**, 7747 (2015).
31. Z. Ni *et al*., *Nat. Energy* **7**, 65–73 (2022).
32. T. Zhu, D. Zheng, M.-N. Rager, T. Pauporté, *Sol. RRL* **4**, 2000348 (2020).
33. T. A. S. Doherty *et al*., *Science* **374**, 1598–1605 (2021).
34. C.-H. Kuan *et al*., *Chem. Eng. J.* **450**, 138037 (2022).
35. Q. Wang *et al*., *Energy Environ. Sci.* **7**, 2359–2365 (2014).
36. R. Tomova, P. Petrova, R. Stoycheva-Topalova, *Phys. Status Solidi. C* **7**, 992–995 (2010).
37. J. Lee *et al*., *Phys. Chem. Chem. Phys.* **18**, 5444–5452 (2016).
38. H. Gao *et al*., *J. Phys. Chem. A* **112**, 9097–9103 (2008).
39. J. Peng *et al*., *Science* **371**, 390–395 (2021).
40. S.-H. Turren-Cruz, A. Hagfeldt, M. Saliba, *Science* **362**, 449–453 (2018).
41. Q. Wang, Q. Dong, T. Li, A. Gruverman, J. Huang, *Adv. Mater.* **28**, 6734–6739 (2016).
42. W. Peng *et al*., *Science* **379**, 683–690 (2023).
43. A. Becheri, M. Dürr, P. Lo Nostro, P. Baglioni, *J. Nanopart. Res.* **10**, 679–689 (2008).
44. Y. Li *et al*., *Nat. Commun.* **12**, 5419 (2021).
45. Y. Lin *et al*., *Nat. Commun.* **9**, 4981 (2018).
46. X. Dai *et al*., *PRX Energy* **1**, 013004 (2022).

ACKNOWLEDGMENTS

Funding: This material is based upon work supported by the US Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office award no. DE-EE0009520. The research at Perotech, Inc., is supported by the Office of Naval Research under award no. N6833522C0122. The work at the University of Toledo was supported by the Center for Hybrid Organic Inorganic Semiconductors for Energy (CHOISE), an Energy Frontier Research Center funded by the Office of Basic Energy Sciences, Office of Science, within the US Department of Energy. DFT calculations used resources of the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract no. DE-AC02-05CH11231 using NERSC award BES-ERCAP0023945. The views expressed herein do not necessarily represent the views of the US Department of Energy or the United States government. **Author contributions:** C.F. and J.H. conceived the idea. C.F. fabricated and characterized perovskite films and devices. M.W. performed the GIXRD measurement and acquired the SEM images. N.L. evaluated several LCMs. X.W. and Y.Y. carried out the DFT calculations. H.G. carried out partial module encapsulation. Z.Z., B.C., and W.X. optimized the laser scribing process. Z.N. carried out the tDOS and DLCP analyses. H.J. carried out the MPP tracker (MPPT) analysis of light stability. Z.S. carried out the PL mapping of the samples. C.F. and J.H. wrote the manuscript, and all authors commented on the manuscript. **Competing interests:** C.F. and J.H. are authors of a provisional patent based on this manuscript. J.H. has disclosed a financial interest with Perotech, Inc. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade9463
Materials and Methods
Figs. S1 to S28
Table S1
References

Submitted 18 September 2022; resubmitted 9 March 2023
Accepted 27 April 2023
10.1126/science.ade9463

OCEAN MICROBIOLOGY

Controlled motility in the cyanobacterium *Trichodesmium* regulates aggregate architecture

Ulrike Pfreundt[1]†, Jonasz Słomka[1]†, Giulia Schneider[1], Anupam Sengupta[2], Francesco Carrara[1], Vicente Fernandez[1], Martin Ackermann[3,4], Roman Stocker[1]*

The ocean's nitrogen is largely fixed by cyanobacteria, including *Trichodesmium*, which forms aggregates comprising hundreds of filaments arranged in organized architectures. Aggregates often form upon exposure to stress and have ecological and biophysical characteristics that differ from those of single filaments. Here, we report that *Trichodesmium* aggregates can rapidly modulate their shape, responding within minutes to changes in environmental conditions. Combining video microscopy and mathematical modeling, we discovered that this reorganization is mediated by "smart reversals" wherein gliding filaments reverse when their overlap with other filaments diminishes. By regulating smart reversals, filaments control aggregate architecture without central coordination. We propose that the modulation of gliding motility at the single-filament level is a determinant of *Trichodesmium*'s aggregation behavior and ultimately of its biogeochemical role in the ocean.

The input of new nitrogen into the ocean and thus global primary productivity is driven in large part by nitrogen fixation by only a few types of cyanobacteria (*1*), one of them being the genus *Trichodesmium* (*2*, *3*). Oceanographers have been puzzled by *Trichodesmium's* dual occurrence as either single multicellular filaments or aggregates comprising hundreds of filaments (*4*–*6*). Both often occur in the same water and are found in all tropical and subtropical oceanic ecosystems (*6*, *7*) and across different *Trichodesmium* species (*8*). In laboratory conditions, aggregation of *Trichodesmium* filaments typically occurs in the postexponential growth phase (*9*–*11*). It can also be induced by depleting iron or phosphate (*12*), reducing salinity (*11*), or changing growth medium or irradiance (*13*). This suggests that aggregation is a general stress response. Compared with filaments, aggregates perform different ecological and metabolic functions (*14*–*21*), partially due to the microbiomes they host (*22*, *23*). Aggregates create different microenvironments (*24*, *25*) that may be transiently suboxic and thus reduce nitrogen and carbon fixation compared with single filaments (*26*). However, through behaviors inaccessible to single filaments, aggregates have better access to other limiting resources. They can capture iron-loaded dust (*14*, *18*, *27*), which may help aggregates engage in iron-intense nitrogen and carbon fixation simultaneously, a hallmark of *Trichodesmium* ecology (*28*). Aggregates may be able to scavenge phosphate from the depths by rapid vertical migration (*15*, *29*), and *Trichodesmium* aggregates at depths below 170 m have been observed to fix nitrogen at similar rates as aggregates in the surface ocean (*30*). Vertical migration is likely also involved in the formation of surface blooms (*31*) spanning tens of thousands of square kilometers (*32*–*34*). However, despite the notable ecology of *Trichodesmium* aggregates little is known about the mechanisms by which filaments give rise to aggregates, how aggregate architecture is controlled, and on what timescale this control occurs. We report that *Trichodesmium* aggregates are active filament assemblies capable of rapid structural changes in response to changes in environmental conditions and we show that this active reshaping of aggregates is realized by changes in the motility of individual filaments.

[1]Institute of Environmental Engineering, Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zürich, Switzerland. [2]Physics of Living Matter, Department of Physics and Materials Science, University of Luxembourg, Luxembourg City, Grand Duchy of Luxembourg. [3]Department of Environmental Systems Sciences, Microbial Systems Ecology Group, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zürich, Switzerland. [4]Department of Environmental Microbiology, Eawag: Swiss Federal Institute of Aquatic Sciences, Dübendorf, Switzerland.
***Corresponding author. Email: romanstocker@ethz.ch**
†These authors contributed equally to this work.

Changes in light exposure induce rapid reshaping of aggregates

We analyzed the effect of changing light conditions on aggregates of *Trichodesmium erythraeum* IMS101, a strain frequently associated with open-ocean surface blooms (*35*). Changes in light intensity are an ecologically relevant cue, occurring in the surface ocean due to cloud movement and can induce cellular stress by generating reactive oxygen species (ROS) (*36*). We exposed naturally aggregating *T. erythraeum* cultures in the late stationary growth phase to sudden changes in light intensity (light switches) (Fig. 1) and quantified the size and density of individual aggregates over ~30 hours through time-lapse imaging.

Aggregates responded rapidly to light switches (movie S1 and Fig. 1). After a Light OFF switch, aggregates expanded (Fig. 1, A and B)—the same set of filaments in the aggregate rearranged into a looser and larger aggregate—causing the average density to decrease by [illegible] 10% (mean ± SEM, n = 18) over 24 min (Fig. 1, C and E, and movie S1). Conversely, after a Light ON switch aggregates tightened causing the average density to increase by 26 ± 17% (mean ± SEM, n = 30) over 65 min (Fig. 1, D and F, and movie S2). The pure circadian onset of the night phase (blue horizontal bar, Fig. 1D) but with the light kept on did not lead to aggregate loosening. Only when the light was actually turned off did the aggregates loosen (at 10 hours, Fig. 1D). Although this does not exclude circadian control of aggregation behavior, it shows that the observed responses were inducible by light changes irrespective of the time of day. Responses were rapid, starting to be visible within 0 to 8 min of a switch (Fig. 1, E and F) and were transient and reversible, i.e., aggregates returned to approximately their previous configuration after cessation of the light perturbation. A UV-killed control culture did not form aggregates (movie S3). The structural configuration of an aggregate can thus change over timescales of minutes, suggesting that active movement is implicated in aggregate loosening and tightening. These experiments also revealed that puff-like aggregates often formed sequentially: individual filaments first aggregated into tight bundles (tufts), which in turn encountered one another and rearranged into puffs (movie S4). This process was also observed to occur in reverse, whereby a puff disintegrated into several tufts (movie S4).

Filament–filament interactions reveal a motility mechanism to modulate aggregate shape

Trichodesmium filaments can glide on surfaces (*12*, *18*, *37*) (movies S5 and S6) through an as-yet unknown locomotion mechanism. We discovered that filaments can also glide on each other without the presence of another surface, in experiments in which we suspended a filament in liquid using a micropipette and brought a second filament into contact with it by fluid flow (movie S7). Upon meeting, the two filaments started gliding against each other without any contact to a solid surface (except the pipette tip). We next characterize the motility of individual filaments and filament pairs and show that filament–filament gliding is a component—although by itself not sufficient—of aggregate reshaping.

Individual *Trichodesmium* filaments move in a series of nearly straight, fast "runs" interrupted by periods of slower motion, which we term reorientations (Fig. 2A). Analyzing the tracks of hundreds of individual filaments (Materials and Methods, Supplementary Text Section I) on the glass surface of a microfluidic chamber (movie S8), we found that filaments glide at 34.1 ± 31.9 μm min^{-1} (mean ± SD, n = 776), primarily along their long axis (fig. S1). Run times exhibit a broad distribution

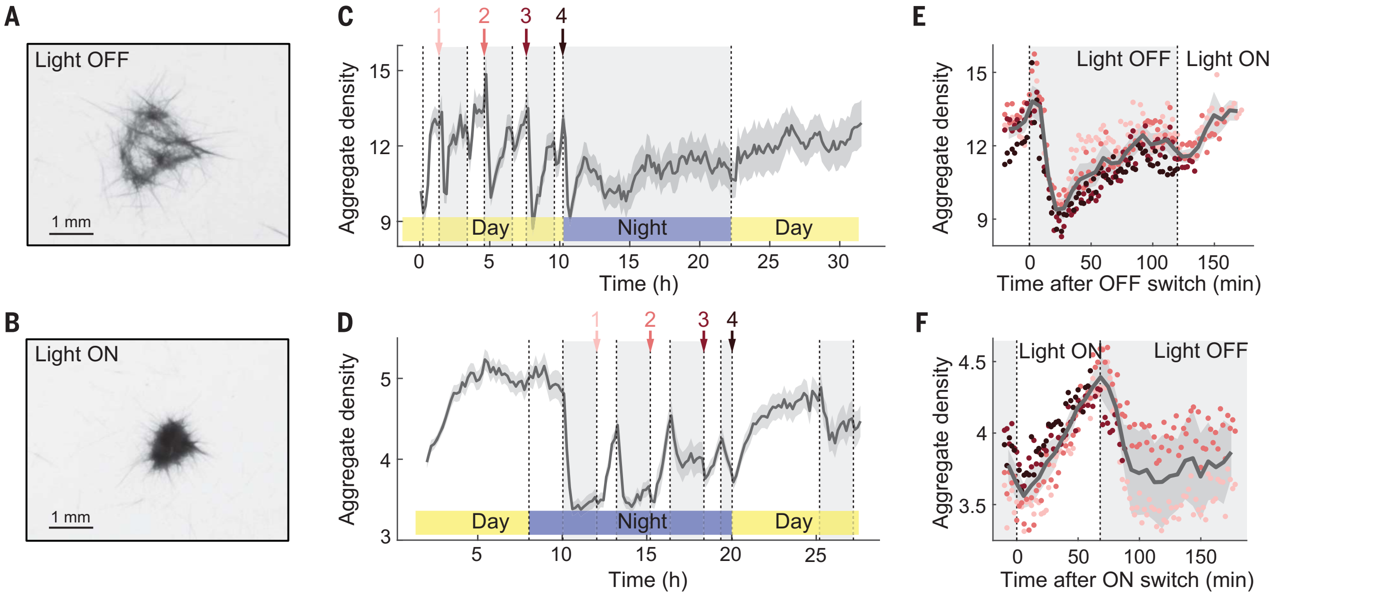


Fig. 1. *Trichodesmium* aggregate density changes rapidly upon switches in light intensity. (**A** and **B**) A puff aggregate in its loose form 15 min after a Light OFF switch (A) and in its dense form in 80 μmol quanta m^{-2} s^{-1} light (B). (**C** and **D**) Time series of mean aggregate density (i.e., the density of filaments within aggregates, quantified as the mean pixel intensity of individual aggregates) in light-switching experiments, with four Light OFF switches (numbered arrows) during the day (C) and four Light ON switches (numbered arrows) during the night (D). Switch 4 corresponds to the normal onset of the next night/day, respectively. Light conditions were 80 μmol quanta m^{-2} s^{-1} in all cases. (**E** and **F**) Overlays of the four Light OFF switches (E) from (C) and the four Light ON switches (F) from (D), color-coded from pink (switch 1) to dark red (switch 4), with each data point representing the mean aggregate density of all aggregates in the imaging frame (16 to 19 aggregates in (E) and 26 to 31 aggregates in (F). In (C to F), aggregate density (arbitrary units) was computed as the mean pixel intensity per aggregate. Note the different scales in (C and D) and (E and F) resulting from the fact that aggregate density is measured as pixel intensity and thus not directly comparable among experiments (Materials and Methods). The gray curves show a 10-min moving average and the gray shading the 95% confidence interval of the mean. We note that the relation between aggregate density and mean pixel intensity may be nonlinear (because intensity saturates at high aggregate density), however it is monotonic (i.e., higher filament density implies higher mean pixel intensity) and thus pixel intensity represents a useful and convenient metric for aggregate density.

(Fig. 2B) with a mean of 5.4 ± 8.2 min (mean ± SD, n = 2416). After a reorientation, filaments continue moving in the same direction or reverse, with similar probability (57 ± 1% versus 43 ± 1%, respectively; mean ± SEM, n = 1518, Fig. 2C). During a reversal, the leading end becomes the trailing end (movie S9), as also observed previously (*12*). Gliding motility allows for fast filament movement but filaments lacking the ability to reverse would quickly glide off a nascent aggregate. We thus inferred that reversals are key to aggregation. However, randomly occurring reversals would still cause an aggregate to disintegrate by filaments gliding off it. Consequently, we hypothesized that reversals are triggered by filament–filament contact, thereby allowing an aggregate to reshape without disintegrating.

In aggregating cultures, filaments within pairs had shorter run lengths and thus reversed more frequently than individual filaments. We induced aggregation in mid to late exponential *T. erythraeum* cultures by addition of menadione (Fig. 2D and movie S10), which causes cellular stress through ROS (*38*). Menadione-induced aggregation is easier to control and thus easier to replicate than relying on post-exponential cultures, in which the exact trigger for aggregation is unknown. We analyzed individual filaments (n = 643) and pairs of filaments that, upon encounter, aligned and glided on each other (n = 59). For both categories, we computed filament reversal frequency as well as different gliding motility statistics that might explain changes in the reversal frequency, including run time, run length, reversal probability, and overall activity (i.e., the fraction of time spent in run mode) (table S1). We found that filaments in pairs moved a smaller fraction of their length during a run, 35% less than single filaments (0.41 ± 0.85 versus 0.63 ± 1.52, respectively; mean ± SD; $n_{\mathrm{runs,pairs}}$ = 377, $n_{\mathrm{runs,single}}$ = 1964; P < 0.001, two-sample one-sided Kolmogorov–Smirnov test) (Fig. 2E). Additionally, filaments in pairs spent more time in run mode (table S1), which induced more reorientation events. This higher activity together with shorter run lengths resulted in a reversal frequency that was nearly twice as high in pairs than that in single filaments (0.034 ± 0.040 min^{-1} versus 0.019 ± 0.031 min^{-1}, respectively; weighted mean ± weighted SD; n_{pairs} = 118, n_{single} = 643; weight = track length; Supplementary Text Section II). The fact that a filament in contact with another filament has a markedly higher reversal frequency than single filaments from the same culture suggests that filaments can respond to each other.

To understand how a filament responds to contact with another filament, we measured "lack-of-overlap" of filament pairs. We define lack-of-overlap as the fraction of the shorter filament's length that does not overlap with the longer filament (Fig. 2F). By tracking both filaments in a pair (n = 59 pairs, as in Fig. 2E), we measured their lack-of-overlap as a function of time (Fig. 2G) to quantify the distribution of lack-of-overlap values. If filaments merely increased their reversal frequency upon contact, the lack-of-overlap would be uniformly distributed (Fig. 2H, Supplementary Text Section III). By contrast, we found that the lack-of-overlap was strongly skewed toward small values ($P < 10^{-6}$, one-sample two-sided Kolmogorov–Smirnov test, n = 53, Fig. 2H). For example, instances of short filaments projecting by less than 50% of their length ($|\text{lack-of-overlap}| < 0.5$) occurred nearly six times more often than filaments projecting by more than 50% ($|\text{lack-of-overlap}| < 0.5$) (Fig. 2H). Thus, the observed increase in reversal frequency upon encounter (Fig. 2E) is not by itself sufficient to describe the behavior

of filaments in pairs. Rather, the prevalence of small lack-of-overlap values suggests that reversals of filaments in pairs are not random but are informed by a sensory cue linked to filament overlap. Thus, reversals are "smart" rather than random. We propose that this behavior is a form of thigmotaxis (*39*, *40*), the change in motility of organisms in response to contact stimuli. In the following, we refer to changes in smart reversals as thigmotaxis, as they are linked to contact between filaments, yet we highlight that the underlying molecular mechanism remains unknown. Such smart reversals are a simple yet effective mechanism for filaments to remain together and rearrange upon encounter. We highlight that filament pairs in the exponentially growing control can also perform smart reversals, however, such pairs disintegrate 2.5 times faster than in the menadione treatment, because filaments in the control perform weaker smart reversals than those treated by menadione, and furthermore because they are faster and shorter (Supplementary Text Section IV; figs. S2 and S3). This faster disintegration rate prevents the formation of stable aggregates in the control.

Mathematical model of smart reversals

To understand the effects on aggregation of different components of the reversal behavior of filaments, we developed an individual-based model of *Trichodesmium* that represents filaments as motile thigmotactic rods capable of sensing overlap with each other and of reversing when they detect that the overlap is decreasing (Fig. 3). Filaments are represented as highly elongated rods (160 μm × 8 μm), consisting of 20 cells that can glide (speed = 32 μm min^{-1}) on surfaces and on other filaments, and adhere to each other upon contact, weakly enough to allow relative gliding (fig. S4, Supplementary Text Section V). All filaments reverse randomly (rate = 32 μm min^{-1} / 160 μm = 0.2 min^{-1}) independently of cues, in line with observations that individual filaments travel approximately their length before reversing (Fig. 2E). Each filament monitors its overlap with filaments in contact with it, where the overlap is defined as the number of neighboring cells on the adjacent filaments a given cell touches, averaged over the filament ("mean cell coordination number"). When a model filament senses a temporal decrease in its overlap, it reverses its gliding direction after a mean response time τ, with smaller values of τ representing greater thigmotaxis strength, i.e., a stronger tendency of filaments to stay together. Below, we vary the response timescale τ in the model to determine the effect of thigmotaxis strength on aggregation.

Our model predicts that thigmotactic filaments form organized (highly overlapping) aggregates on a surface (Fig. 3A and movie S11), akin to the dynamic aggregates observed in *T. erythraeum* exposed to menadione (Fig. 2D and movie S10). To quantify aggregation in model simulations, we compute the average overlap over all filaments in the system as a function of time (Fig. 3B). We focus on the three values of thigmotaxis strength, representing filaments that, after detecting a decrease in overlap (which takes ~10 s), (i) reverse after a short delay (τ = 0.75 s, strong thigmotaxis); (ii) reverse after an intermediate delay (τ = 12 s, medium thigmotaxis); and (iii) reverse after a long delay (τ = 75 s, weak thigmotaxis). Equivalently, for filaments in a pair moving in opposite directions and risking separation, the three model conditions of strong, intermediate,

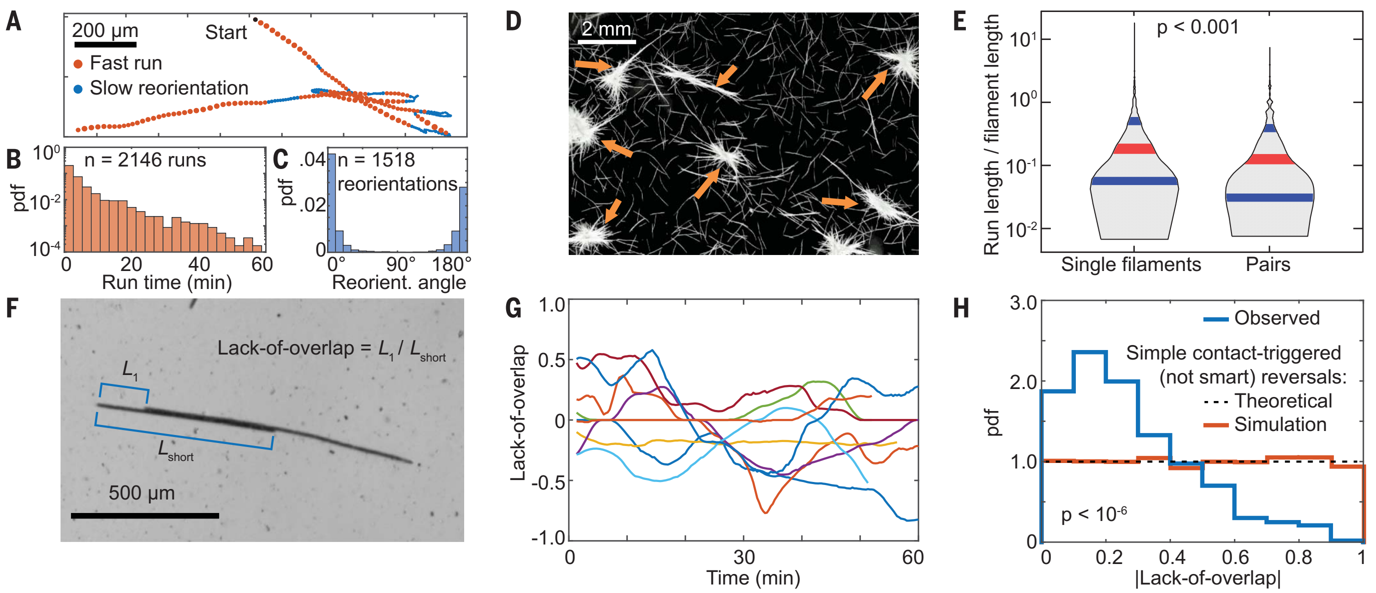


Fig. 2. Controlled reversals characterize *Trichodesmium* aggregation. (**A**) Trajectory of a filament gliding on a surface, segmented into runs (orange) and reorientations (blue). Dots mark the position of the centroid, with size proportional to instantaneous speed. (**B**) Run times exhibit a broad distribution. (**C**) Reorientation angles exhibit a nearly equal repartition between values close to 0° (continuation in the same direction) and values close to 180° (reversals). (**D**) Midexponential *Trichodesmium* filaments exposed to menadione, which induces oxidative stress, form aggregates (orange arrows). (**E**) The run length (normalized by filament length) of menadione-exposed *Trichodesmium* is markedly shorter (35% decrease in the mean) for filaments in pairs (two filaments in contact and aligned) than for single filaments [n = 12 independent experiments; $P < 0.001$, two-sample one-sided Kolmogorov–Smirnov test (n_{single} = 1964, n_{pairs} = 377)]. The gray area shows the distributions as violin plots. The red line shows the median and the blue lines show the 25th and 75th percentiles. (**F**) Filament pair illustrating the lack-of-overlap, defined as the fraction of the shorter filament length not overlapping with the longer filament. Lack-of-overlap is positive at one end of the filament pair and negative at the other end. (**G**) Lack-of-overlap as a function of time for nine filament pairs, illustrating the dynamic rearrangement between filaments. (**H**) The lack-of-overlap distribution of menadione-exposed filament pairs is strongly skewed toward small lack-of-overlap values compared with the uniform distribution expected if filaments were to simply increase reversal frequency upon contact ($P < 10^{-6}$, one-sample two-sided Kolmogorov–Smirnov test, n = 53). This is consistent with the hypothesis that *Trichodesmium* performs "smart reversals", whereby reversals are triggered by an increase in lack-of-overlap as filaments glide upon each other (see text).

and weak thigmotaxis correspond to traveling approximately 4, 15, and 55% of a filament length before performing a smart reversal. For strong thigmotaxis (τ = 0.75 s), smart reversals lead to the formation of tight aggregates, characterized by high overlap, by preventing filaments from escaping from aggregates (Fig. 3, A and B): when filaments sense a decrease in their overlap with other filaments in the aggregate they rapidly trigger a reversal that on average redirects them toward the aggregate. For strong thigmotaxis, aggregation is robust against filament length variability (fig. S5). Medium (τ = 12 s) or weak (τ = 75 s) thigmotaxis strengths result in looser aggregates containing fewer filaments or almost no aggregates, respectively (Fig. 3B). Identifying filament pairs in these simulations shows that the response time τ controls the skewness of the lack-of-overlap distribution (Fig. 3C), thus linking the response time τ—a core model parameter characterizing smart reversals—with the lack-of-overlap distribution observed in experiments (Fig. 2H). When τ is small (strong thigmotaxis), the lack-of-overlap is skewed toward small values (Fig. 3C), in line with our observations (Fig. 2H). As τ increases (medium and weak thigmotaxis) the lack-of-overlap distribution becomes more uniform (Fig. 3C).

The response timescale τ can be approximately measured in our experiments from the rate of disintegration of filament pairs (Supplementary Text Section VI). These measurements indicate that filaments in menadione correspond to medium-to-strong thigmotaxis in the model, whereas filaments in the control correspond to weak-to-medium thigmotaxis in the model (fig. S5). Additionally, our data show that slower or longer filaments perform stronger thigmotaxis than faster or shorter ones, respectively (Supplementary Text Section VII, table S2).

We also modeled aggregate formation in 3D to show that smart reversals are key to the formation of organized aggregates in a liquid suspension (i.e., away from solid surfaces), such as the marine habitat of *Trichodesmium*. This model represents sequential random arrivals of randomly oriented filaments onto a single seed filament, with which they align upon encounter (Fig. 3D), mimicking, for example, encounters and alignment driven by ocean turbulence (fig. S6, Supplementary Text Sections VIII and IX). We find that strong thigmotaxis converts successive random encounters into an organized aggregate characterized by large overlap (blue curve in Fig. 3D and movie S12). Such large overlap (>5), which approaches values characterizing random packing of spheres (*41*), indicates that a random cell within the aggregate has typically five or more neighboring cells, not including cells on the same filament. Switching off smart reversals after an organized aggregate has formed (thus leaving only random reversals) leads to rapid disintegration of the aggregate (purple curve in Fig. 3D). Similarly, random reversals alone do not lead to stable aggregates, with filaments only forming small, transient aggregates that continuously disintegrate [for both high (2 min^{-1}) and low (0.4 min^{-1}) rates of random reversals, pink and cyan in Fig. 3D]. Additionally, simple adhesion upon encounters, without rearrangement, would result in fractal aggregate morphologies (*42*) (yellow curve in Fig. 3D and movie S13). Lastly, if filaments stuck to each other and aligned in flow but were otherwise nonmotile they would form elongated chains (red curve in Fig. 3D and movie S14). Both fractals and chains could emerge in scenarios where filaments were nonmotile and sticky. Only strong thigmotaxis produces organized aggregates characterized by large overlap (approximately 5 to 6), whereas all other mechanisms considered above resulted in values of overlap that were at least 50% smaller.

Smart reversals can also explain the rapid, dynamic tightening and loosening of aggregates that we observed upon switches in light intensity (Fig. 1). Modeling the response to an increase and decrease in light intensity as strong and medium thigmotaxis, respectively, reproduces the observed tightening and loosening of aggregates (Fig. 3E). Specifically, the loosening of a compact model aggregate by

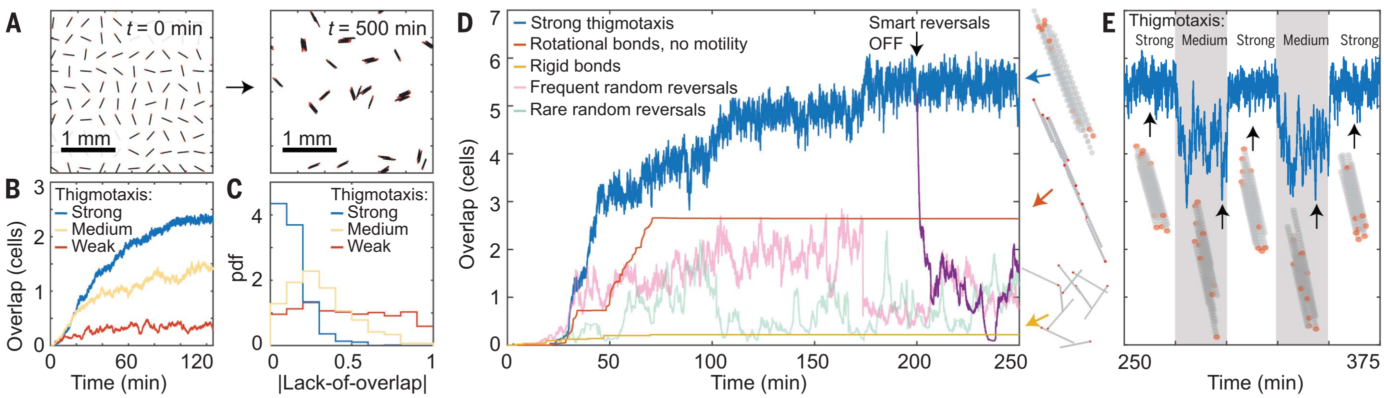


Fig. 3. Organized aggregate architecture in 3D emerges from smart reversals. (**A**) In our 2D model of aggregation on a surface, filaments performing smart reversals (τ = 0.75 s, *n* = 100), initially separated and oriented randomly on the surface, form aggregates of oscillating filaments (movie S11), akin to those observed experimentally (Fig. 2D). (**B**) The time course of filament overlap (mean cell coordination number) on a surface for filaments performing smart reversals with different response times (τ = 0.75 s, blue; τ = 12 s, yellow; τ = 75 s, red) shows that tuning the response time directly controls the strength of thigmotaxis: small values of τ realize strong thigmotaxis characterized by the emergence of highly overlapping aggregates. Displayed are averages over 6 runs for simulations with 20 filaments. (**C**) Distribution of the lack-of-overlap of filament pairs identified in 2D simulations (at least *n* = 6) for different thigmotaxis strengths, showing that thigmotaxis skews the lack-of-overlap toward small values, consistent with the experimental observations (Fig. 2H). (**D**) Overlap as a function of time predicted by 3D simulations of 10 filaments in suspension. Aggregates form from an initial seed filament that collects other filaments through random encounters. The effect of three possible mechanisms following filament encounter is shown: (i) filaments form rigid bonds and do not rearrange, which creates fractal-like aggregates (yellow); (ii) filaments are not motile and align under external torques (e.g., turbulence-induced), which creates elongated chains (red); or (iii) motile filaments perform smart reversals with strong thigmotaxis (τ = 0.75 s), which creates compact aggregates (blue). Switching off smart reversals (while retaining random reversals) leads to rapid disintegration of aggregates (purple). Finally, random reversals do not create stable aggregates (pink and cyan). (**E**) Continuation of the model simulation for the aggregate of filaments executing strong thigmotaxis [blue curve in panel (D)], when the thigmotaxis strength is modulated between strong (τ = 0.75 s) and medium (τ = 12 s). The resulting tightening and loosening of the aggregate reproduces the response of aggregates exposed to light switches (Fig. 1).

~30% (measured as the change in overlap in the model; Fig. 3E and movie S15) is akin to the change in aggregate density (a proxy for total overlap) observed experimentally (Fig. 1 and movie S16). Thus, a change in single-filament behavior—modeled as a change in the response time τ between detecting a decrease in overlap and reversing—can explain the rapid reshaping of aggregates in response to light switches.

Smart reversals represent an effective mechanism for filaments to remain together after an encounter. These encounters occur differently on a surface versus in the ocean. On a surface, as in most of our experiments, gliding can drive encounters, which occur on the timescale of an hour (fig. S3E). By contrast, in the three-dimensional (3D) water column without surfaces to glide on, encounters are most likely driven by turbulence and buoyancy. The typical encounter timescale between buoyant (density offset 50 kg m^{-3}) individual filaments is <20 hours (*43*) for a turbulent intensity characteristic of the ocean surface layer (10^{-6} W kg^{-1}) at an organism concentration representing bloom conditions (10^7 filaments m^{-3}) (*44*). Consequently, the encounter timescale in the ocean is short enough to convert individual filaments into aggregates within several days, even at tenfold lower concentrations, as we demonstrated using a coagulation model that accounts for aggregate-aggregate encounters (fig. S7, Supplementary Text Section X). By contrast, because *Trichodesmium*'s doubling time is a few days (*45–47*), it would take several weeks for a single filament to grow into a sizeable aggregate, suggesting that encounters are likely faster than growth in driving the formation of aggregates. Still, an encounter timescale of many hours is much longer than the tens of minutes required for a filament to travel its length. Smart reversals protect the nascent aggregate from disintegrating: if filaments lose each other it may be hours before a new filament is encountered. Additionally, turbulence in the ocean surface layer is unlikely to separate a pair of filaments stuck to each other as it exerts shear forces on the order of tens of picoNewtons (Supplementary Text Section X), much smaller than filament adhesive forces, likely mediated by cellular appendages, which are in the range of tens of nanoNewtons (*12*).

Discussion

In our experiments, puff-like aggregates are often formed by merging tufts, which in turn could form from disintegrating puffs (movie S4). This sequential formation of puffs is corroborated by dilution experiments, in which only tufts formed (fig. S8). In the ocean, *Trichodesmium* filaments are dilute (10^4 to 10^6 m^{-3}) (*3*, *6*), separated from each other by many filament lengths, corresponding to the conditions of our model of aggregate formation in 3D: puffs may then emerge from encounters between tufts. Puffs may also form through other mechanisms (e.g., aggregation on dust particles) (*14*, *48*) and some strains may form only puffs or tufts (*8*, *23*). Our work, however, indicates that these two morphologies are less static than currently assumed and can reshape into one another. This might be broadly applicable as most *Trichodesmium* clades (except Clade IV) occur as both tufts and puffs (*23*). *Trichodesmium* aggregates can constitute different clades (*23*), corroborating the idea of encounter-driven aggregation. Whether clades can tune smart reversals to avoid other strains such as nondiazotrophic species (*49*) and thus manipulate aggregate composition remains an open question.

Aggregates have been suggested to form through increased stickiness, mediated by appendages (*12*) or exopolymers (*50*). However, our model shows that adhesion alone does not lead to organized aggregates. Although a role for motility in aggregation has been suggested (*12*, *18*, *37*), uncoordinated motility would lead to aggregate disintegration and cannot account for reversible aggregate tightening and loosening. Our experiments instead indicate that *T. erythraeum* adjusts its motility in response to a sensory cue, related to the overlap with other filaments. We propose that this response is triggered by surface contact, similar to C-signaling in myxobacteria (*51*). We find that intrafilament cell–cell autoinduction could plausibly trigger reversals (fig. S9) but more work is needed to unravel molecular mechanisms.

Our work suggests that smart reversals allow *Trichodesmium* to react to environmental changes as transient as shifting cloud cover. Aggregates may loosen to gain exposure to light or tighten to self-shade (in this context, it would be interesting to quantify the response of menadione-treated cultures to changes in light). The transient response we observed suggests that it represents the initial, fastest adaptation to environmental changes ahead of more drastic adaptations (e.g., proteomic reorganization). The rapid changes to aggregate density also modulate exposure to gases and chemicals by the cells and their microbiomes, which may allow them to tune metabolic processes and vertical migration.

Trichodesmium aggregation through motility is an example of an active matter system (*52*, *53*). It bears analogies with the formation of fruiting bodies in *Myxococcus xanthus*, resulting from jamming driven by high cell densities, gliding, and reversals (*54*) and with the behavior of *Pseudomonas aeruginosa*, which avoids jamming by reversing upon contact (*55*). *Trichodesmium* aggregation is different, however, as its concentration in the ocean is too low for jamming (Supplementary Text Section XI). Instead, *Trichodesmium* uses smart reversals to convert encounters into aggregates. Aggregation also occurs in other filamentous cyanobacteria [e.g., *Oscillatoria terebriformis* (*56*), *Nostoc punctiforme* (*57*, *58*), and in multispecies freshwater cyanospheres (*59*, *60*)]. However, in all cases aggregation mechanisms remain unclear.

Previous work has shown that *Trichodesmium* filaments can wiggle to convey captured iron particles to the core of their aggregate (*14*, *18*, *27*). We showed that filament behavior further controls the aggregate's structure itself (fig. S10). Ultimately, the biogeochemical importance of *Trichodesmium*—as a nitrogen fixer, primary producer, and vector of CO_2 sequestration to depth—hinges on the decentralized yet coordinated behavior of individual filaments.

REFERENCES AND NOTES

1. C. Martínez-Pérez *et al.*, *Nat. Microbiol.* **1**, 16163 (2016).
2. E. J. Carpenter, D. G. Capone, in *Nitrogen In The Marine Environment* (Elsevier, 2008), pp. 141–198.
3. R. M. Letelier, D. M. Karl, *Mar. Ecol. Prog. Ser.* **133**, 263–273 (1996).
4. M. Rodier, R. Le Borgne, *Mar. Pollut. Bull.* **61**, 349–359 (2010).
5. I. Bryceson, P. Fay, *Mar. Biol.* **61**, 159–166 (1981).
6. E. J. Carpenter, A. Subramaniam, D. G. Capone, *Deep Sea Res. Part I Oceanogr. Res. Pap.* **51**, 173–203 (2004).
7. C. S. Davis, D. J. McGillicuddy Jr., *Science* **312**, 1517–1520 (2006).
8. A. M. Hynes, E. A. Webb, S. C. Doney, J. B. Waterbury, *J. Phycol.* **48**, 196–210 (2012).
9. P. R. F. Bell *et al.*, *Hydrobiologia* **532**, 9–21 (2005).
10. L. Prufert-Bebout, H. W. Paerl, C. Lassen, *Appl. Environ. Microbiol.* **59**, 1367–1375 (1993).
11. F.-X. Fu, P. R. F. Bell, *Mar. Ecol. Prog. Ser.* **257**, 69–76 (2003).
12. Y. Tzubari, L. Magnezi, A. Be'er, I. Berman-Frank, *ISME J.* **12**, 1682–1693 (2018).
13. Y.-B. Chen, J. P. Zehr, M. Mellon, *J. Phycol.* **32**, 916–923 (1996).
14. M. Rubin, I. Berman-Frank, Y. Shaked, *Nat. Geosci.* **4**, 529–534 (2011).
15. A. E. White, Y. H. Spitz, R. M. Letelier, *Mar. Ecol. Prog. Ser.* **323**, 35–45 (2006).
16. K. R. Frischkorn, M. Rouco, B. A. S. Van Mooy, S. T. Dyhrman, *ISME J.* **11**, 2090–2101 (2017).
17. K. R. Frischkorn, S. T. Haley, S. T. Dyhrman, *ISME J.* **12**, 997–1007 (2018).
18. J. G. Rueter, D. A. Hutchins, R. W. Smith, N. L. Unsworth, in *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs*, E. J. Carpenter, D. G. Capone, J. G. Rueter, Eds. (Springer, 1992), pp. 289–306.
19. K. R. Frischkorn, S. T. Haley, S. T. Dyhrman, *Front. Microbiol.* **10**, 330 (2019).
20. M. D. Lee *et al.*, *ISME J.* **11**, 1813–1824 (2017).
21. B. A. S. Van Mooy *et al.*, *ISME J.* **6**, 422–429 (2012).
22. S. Basu, M. Gledhill, D. de Beer, S. G. Prabhu Matondkar, Y. Shaked, *Commun. Biol.* **2**, 284 (2019).
23. M. Rouco, S. T. Haley, S. T. Dyhrman, *Environ. Microbiol.* **18**, 5151–5160 (2016).
24. M. Eichner, S. Basu, M. Gledhill, D. de Beer, Y. Shaked, *Front. Microbiol.* **10**, 1565 (2019).
25. M. J. Eichner *et al.*, *ISME J.* **11**, 1305–1317 (2017).
26. M. Eichner *et al.*, *New Phytol.* **222**, 852–863 (2019).
27. N. Kessler *et al.*, *ISME J.* **14**, 91–103 (2020).
28. N. A. Held *et al.*, *Nat. Microbiol.* **7**, 300–311 (2022).
29. T. A. Villareal, E. J. Carpenter, *Microb. Ecol.* **45**, 1–10 (2003).
30. M. Benavides *et al.*, *ISME J.* **16**, 2398–2405 (2022).
31. M. Rodier, R. Le Borgne, *J. Exp. Mar. Biol. Ecol.* **358**, 20–32 (2008).
32. D. G. Capone *et al.*, *Mar. Ecol. Prog. Ser.* **172**, 281–292 (1998).
33. D. M. Karl, R. Letelier, D. V. Hebel, D. F. Bird, C. D. Winn, in *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs*, E. J. Carpenter, D. G. Capone, J. G. Rueter, Eds. (Springer, 1992), pp. 219–237.
34. E. J. Carpenter, D. G. Capone, in *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs*, E. J. Carpenter, D. G. Capone, J. G. Rueter, Eds. (Springer, 1992), pp. 211–217.
35. K. M. Orcutt *et al.*, *Appl. Environ. Microbiol.* **68**, 2236–2245 (2002).
36. N. Blot *et al.*, *Plant Physiol.* **156**, 1934–1954 (2011).
37. H. W. Paerl, B. M. Bebout, in *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs*, E. J. Carpenter, D. G. Capone, J. G. Rueter, Eds. (1992), pp. 43–59.
38. X. Kammerscheit, F. Chauvat, C. Cassier-Chauvat, *Front. Microbiol.* **10**, 1899 (2019).
39. S. J. Schnörr, P. J. Steenbergen, M. K. Richardson, D. L. Champagne, *Behav. Brain Res.* **228**, 367–374 (2012).

40. J. Sikora, Z. Baranowski, M. Zajaczkowska, *Experientia* **48**, 789–792 (1992).
41. L. V. Migal, V. G. Bondarev, N. A. Chekanov, T. P. Bondareva, *J. Phys. Conf. Ser.* **1479**, 012097 (2020).
42. J. R. Rothenbuhler, J.-R. Huang, B. A. DiDonna, A. J. Levine, T. G. Mason, *Soft Matter* **5**, 3639–3645 (2009).
43. J.-A. Arguedas-Leiva, J. Słomka, C. C. Lalescu, R. Stocker, M. Wilczek, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2203191119 (2022).
44. W. M. Dunstan, J. Hosford, *Bull. Mar. Sci.* **27**, 824–829 (1977).
45. E. Breitbarth, J. Wohlers, J. Kläs, J. LaRoche, I. Peeken, *Mar. Ecol. Prog. Ser.* **359**, 25–36 (2008).
46. C. M. Holl, J. P. Montoya, *J. Phycol.* **44**, 929–937 (2008).
47. M. R. Mulholland, D. G. Capone, *Mar. Ecol. Prog. Ser.* **188**, 33–49 (1999).
48. N. A. Held *et al*., *ISME Commun.* **1**, 35 (2021).
49. T. O. Delmont, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2112355118 (2021).
50. I. Berman-Frank, G. Rosenberg, O. Levitan, L. Haramaty, X. Mari, *Environ. Microbiol.* **9**, 1415–1422 (2007).
51. D. Kaiser, *Annu. Rev. Microbiol.* **58**, 75–98 (2004).
52. M. Bär, R. Großmann, S. Heidenreich, F. Peruani, *Annu. Rev. Condens. Matter Phys.* **11**, 441–466 (2020).
53. M. C. Marchetti *et al*., *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
54. G. Liu *et al*., *Phys. Rev. Lett.* **122**, 248102 (2019).
55. M. J. Kühn *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2101759118 (2021).
56. R. W. Castenholz, *Nature* **215**, 1285–1286 (1967).
57. A. Wilde, C. W. Mullineaux, *Mol. Microbiol.* **98**, 998–1001 (2015).
58. A. Guljamow *et al*., *Appl. Environ. Microbiol.* **83**, e01510–e01517 (2017).
59. K. Milferstedt *et al*., *Sci. Rep.* **7**, 17944 (2017).
60. S. J. N. Duxbury, S. Raguideau, J. Rosko, K. Cremin, M. Coates, C. Quince, O. S. Soyer, Reproducible spatial structure formation and stable community composition in the cyanosphere predicts metabolic interactions. bioRxiv 2022.12.13.520286 [Preprint] (2022); https://doi.org/10.1101/2022.12.13.520286.
61. J. Slomka, Data Collection for the article 'Controlled motility in the cyanobacterium *Trichodesmium* regulates aggregate architecture', ETH Research Collection (2022); https://doi.org/10.3929/ethz-b-000596515.

ACKNOWLEDGMENTS

We thank W. Hess for *Trichodesmium* cultures, E. Borowiecka for help with the figures, and O. Soyer, R. Naisbit, U. Alcolombri, and J. Keegstra for discussions. **Funding:** We gratefully acknowledge funding from an ETH Zurich Postdoctoral Fellowship to U.P.; from an ETH Zurich Postdoctoral Fellowship and a Swiss National Science Foundation Ambizione Grant (PZ00P2_202188) to J.S.; from an ATTRACT Investigator Grant (A17/MS/11572821/MBRACE) and FNR-CORE Grant (C19/MS/13719464/TOPOFLUME/Sengupta) from the Luxembourg National Research Fund to A.S.; from the Simons Foundation through the Principles of Microbial Ecosystems (PriME) collaboration (grant 542389) to M.A. and (grant 542395) to R.S.; from a Gordon and Betty Moore Foundation Symbiosis in Aquatic Systems Investigator Award (GBMF9197), the Swiss National Science Foundation, National Centre of Competence in Research (NCCR) Microbiomes (51NF40_180575), and a Swiss National Science Foundation research grant (205321_207488) to R.S. **Author contributions:** U.P. and J.S. contributed equally to this work and are listed in alphabetical order. U.P., J.S., A.S., M.A. and R.S. designed research. U.P. and G.S. conducted experiments. J.S. developed mathematical models. U.P., G.S., J.S. and R.S. analyzed data. A.S. provided support with initial experiments and preliminary data analyzes. U.P., J.S., M.A., and R.S. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript, the supplementary material or deposited at ETH Research Collection (*61*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2753
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S3
References (*62*–*77*)
MDAR Reproducibility Checklist
Movies S1 to S16

Submitted 11 October 2022; accepted 25 April 2023
10.1126/science.adf2753

PLANT ECOLOGY

Shifting microbial communities can enhance tree tolerance to changing climates

Cassandra M. Allsup†, Isabelle George†, Richard A. Lankau*

Climate change is pushing species outside of their evolved tolerances. Plant populations must acclimate, adapt, or migrate to avoid extinction. However, because plants associate with diverse microbial communities that shape their phenotypes, shifts in microbial associations may provide an alternative source of climate tolerance. Here, we show that tree seedlings inoculated with microbial communities sourced from drier, warmer, or colder sites displayed higher survival when faced with drought, heat, or cold stress, respectively. Microbially mediated drought tolerance was associated with increased diversity of arbuscular mycorrhizal fungi, whereas cold tolerance was associated with lower fungal richness, likely reflecting a reduced burden of nonadapted fungal taxa. Understanding microbially mediated climate tolerance may enhance our ability to predict and manage the adaptability of forest ecosystems to changing climates.

Climate change is forcing populations outside of their range of climate tolerance, with estimates as high as 16% of species facing extinction in the next century in response (*1*–*3*). When faced with changing climates, species must either migrate to track optimal climates or adapt or acclimate to the new conditions; failing that, they face extinction (*1*, *4*). However, species exist within diverse communities, and interactions among species may provide alternative pathways to climate tolerance (*5*, *6*). In particular, macroscopic organisms exist in constant association with diverse communities of microbes (i.e., their microbiome), which shape their phenotypes and their responses to stresses such as climatic change (*7*, *8*).

Plants have repeatedly evolved relationships with microbes to meet essential needs (*9*, *10*). Rhizosphere and endophytic microbes often enhance plant tolerance to a range of stresses when compared with axenic plants (*7*, *8*), including cold (*11*), heat (*12*), and drought stress (*13*–*15*). Associations with specific microbial taxa can be essential to the adaptation of plant populations to extreme environments (*12*, *16*, *17*), and microbial communities sourced from specific environments can affect plant responses to similar environmental stresses in artificial conditions (*18*–*20*). Almost all tree species form associations with either arbuscular (AM) or ectomycorrhizal (EM) fungi, which represent phylogenetically and functionally distinct symbioses. Mycorrhizal fungi can enhance plant tolerance to many environmental conditions (*7*, *21*, *22*) and are thus prime candidates for microbes responsible for microbially mediated climate tolerance. Such "stress-adapted" microbial communities could influence plant performance by including microbial strains better able to perform under the stress or by excluding detrimental microbial taxa that cannot tolerate the stressful condition. If microbial associations broadly contribute to determining realized plant-climate niches, then our ability to understand current, and predict future, plant distributions may be incomplete. However, the extent to which variation in microbial communities across climatic gradients can lead to enhancement of host plant tolerance to those same gradients is largely untested in natural conditions. Microbial taxa are likely to adapt faster than their host plants, especially trees (*23*, *24*), and may disperse farther (*25*, *26*). This may provide an alternative source of plant community resistance to climate change through microbial adaptation in place or migration through space.

Increased temperatures and drought are predicted to reshape forests through changes in species abundance (*27*), redistribution of species ranges (*28*, *29*), and conversions of forest to other biomes (*30*, *31*), with likely feedbacks to the global climate system (*32*). To investigate how geographic variation in microbial communities could affect tree responses to climate stressors, we first collected microbial communities sourced from 12 locations along gradients of temperature and precipitation. Then, we tested how inoculation with these communities influenced tree seedlings' responses to climate stresses in two locations chosen to represent expanding and trailing range edges, as well as in controlled conditions. At each location, we grew seedlings of several tree genera, including those that associate with either AM or EM fungi, in ambient and reduced rainfall conditions. Across the two sites and the rainfall treatments, this resulted in seedlings experiencing variation in cold, heat, and drought stress. Seedling survival was enhanced by inoculation with microbial communities sourced from areas with climates that best matched the major climate stress facing the seedlings in each field site or experimental treatment. The microbially enhanced tolerance to heat and drought only occurred for trees associating with AM fungi, whereas cold tolerance occurred across mycorrhizal types. These results indicate that soil and root microbial communities can provide an alternative pathway to climate tolerance for forests.

Matching of microbial community source and climate stress–enhanced survival in field conditions

At our northern site (Fig. 1), we planted a mixture of tree species that were either native to the site (but with range centers south of the site) or had native ranges south of the site (table S1). This experiment simulated the process of range expansion, in which populations expand or establish in areas that have become permissive because of climate change. We expected that survival over winter would be a primary climate stress on seedlings at this site. Therefore, we predicted that seedling survival would be maximized when preinoculated with microbial communities sourced from colder areas (microbially induced cold tolerance). Supporting this hypothesis, in ambient rainfall conditions, seedlings inoculated with soil microbes sourced from areas with colder winters had higher 3-year survival rates [phylogenetic generalized linear mixed model (PGLMM), $P = 0.031$; Fig. 2A and table S2]. This effect was not apparent when seedlings were grown under reduced rainfall conditions (PGLMM, $P = 0.823$; Fig. 2B and table S2). The minimum temperature of the microbial-source site was a better predictor of seedling survival than the mean or maximum temperatures of the source site (*33*). The aridity of the microbial source was not significantly associated with seedling survival (table S2). Survival was independently influenced by the initial seedling height (table S2).These results did not differ between seedlings associating with AM or EM fungi (Fig. 2, A and B, and table S3).

To examine how microbial-source climate affected survival across the 3-year experiment, we compared survival across distinct seasons ["growing season" spring-to-fall survival versus "overwintering" fall-to-spring survival (fig. S1)]. During the first summer, survival was highest in seedlings inoculated with microbial communities from warmer locations in both ambient and reduced rainfall conditions (PGLMM, $P = 0.064$ and 0.031, respectively; Fig. 2, C and D, and table S4). However, over the first winter period this pattern reversed; overwinter survival was significantly higher for seedlings inoculated with microbial communities from colder sites (PGLMM, $P < 0.0001$ for ambient conditions, $P = 0.019$ for rainfall-reduced conditions; Fig. 2, C and D, and table S4). During the second summer, survival in the ambient versus rainfall-reduced

Department of Plant Pathology, University of Wisconsin–Madison, Madison, WI, USA.
***Corresponding author. Email: lankau@wisc.edu**
†These authors contributed equally to this work.

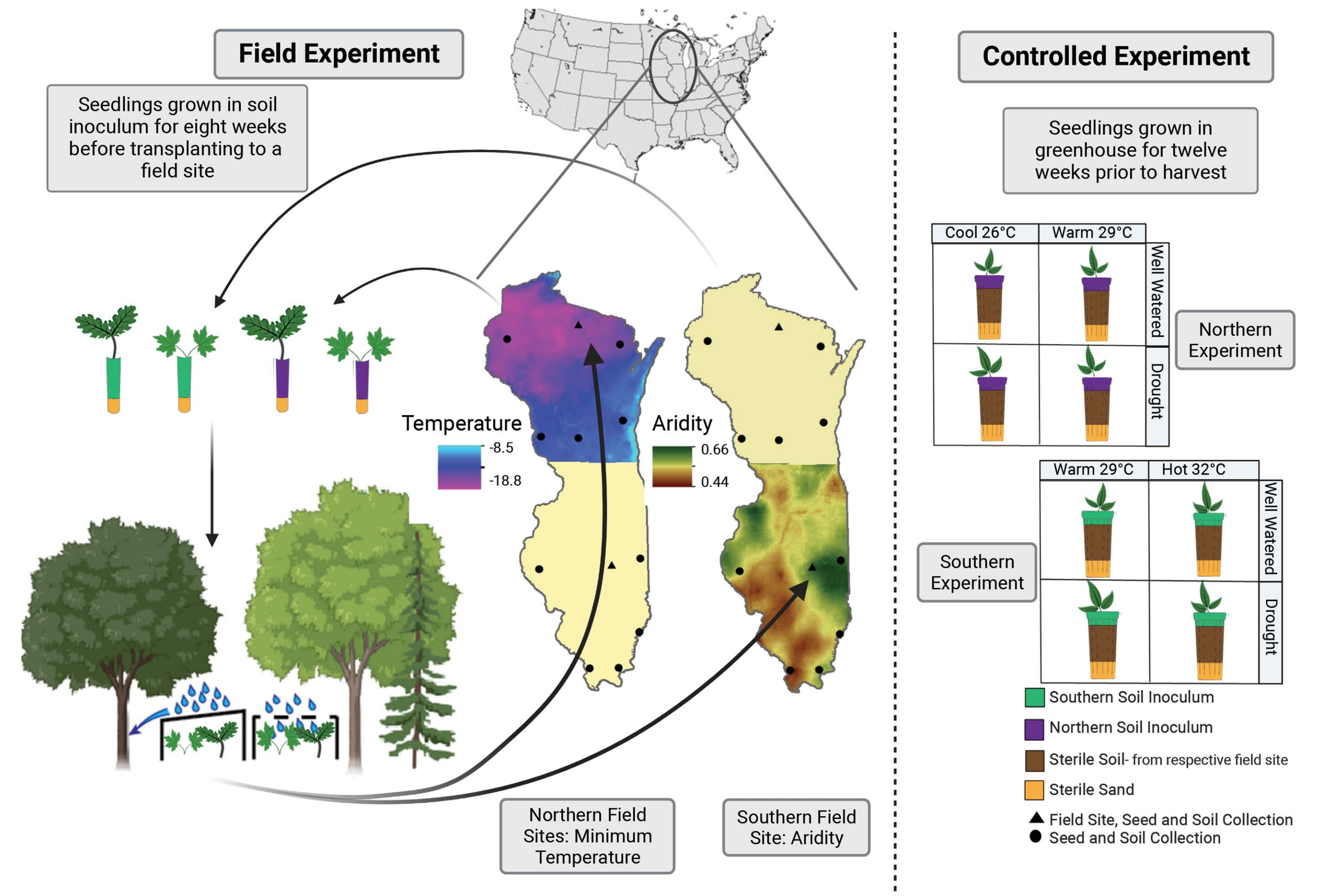


Fig. 1. Experimental design of linked field and greenhouse experiments. Parallel sets of experiments were performed for the northern sites and soils to simulate range expansion and for the southern site and soils to simulate range contraction. (i) Live soil was collected from six sites in each state. (ii) Seedlings were germinated in sterile conditions and then grown for 8 weeks with live field soil prior to transplantation into experimental sites (triangle symbols). (iii) Seedlings were grown under rainfall-reduction or mock (ambient condition) shelters in field sites for 3 years, with survival monitored seasonally, for each of two cohorts (planted in 2018 and 2019). (iv) Live soil from the same sites was used to inoculate pots for a controlled-environment experiment in which pots were filled with sterilized background soil from the field experiment location. (v) Seedlings were grown for 12 weeks in two temperature conditions crossed with two watering frequencies. The greenhouse experiment was repeated three times, once per year for 3 years, with new soil collections for each repetition.

conditions diverged: survival in ambient conditions reinforced the trend seen in the previous winter, whereas survival in the rainfall-reduced conditions counteracted that trend (PGLMM, rainfall treatment × microbial-source minimum temperature interaction, $P = 0.025$; Fig. 2, C and D, and table S4). That is, under ambient rainfall conditions, seedlings faced one major climatic stress—cold winter temperatures—and survival was optimized by preinoculation with microbial communities from colder sites. But in the rainfall-reduced treatment, seedlings faced multiple climatic stresses (winter cold as well as summer drought), and thus no particular microbial community was able to optimize cumulative survival over the 3-year period.

We chose our southern site in central Illinois to reflect heat and drought stress for our suite of temperate tree species. This location is south of the range center for all of our target species' ranges and near the southern range limit for many. Therefore, this location represented dynamics at a species' trailing edge where climates are becoming too warm or dry for population persistence (table S1). Here, we expected that in the ambient conditions, seedlings would be primarily stressed by high summer temperatures, and we hypothesized that seedling survival would be optimized when they were inoculated with microbial communities sourced from warmer sites (i.e., microbially induced heat tolerance). Alternatively, in our rainfall-reduction treatment, we expected that seedlings would be stressed to a greater relative extent by drought and thus hypothesized that seedling survival would be optimized when inoculated with microbial communities sourced from more arid sites (microbially induced drought tolerance). In partial support of our hypotheses, seedling survival trended toward a positive relationship with the maximum temperature of the microbial-source site for seedlings in our ambient rainfall treatment (PGLMM, $P = 0.093$; Fig. 3A and table S5) but not in the rainfall-reduced treatment (PGLMM, $P = 0.952$; Fig. 3A and table S5). Seeding survival was very strongly promoted by preinoculation with microbial communities from more arid sites in our rainfall-reduced treatment (PGLMM, $P = 0.002$; Fig. 3B and table S5) but not in our ambient rainfall treatment (PGLMM, $P = 0.940$; Fig. 3B and table S5). Qualitatively similar results were found when analyzing survival time with Cox regression at both experimental sites (tables S2 and S5).

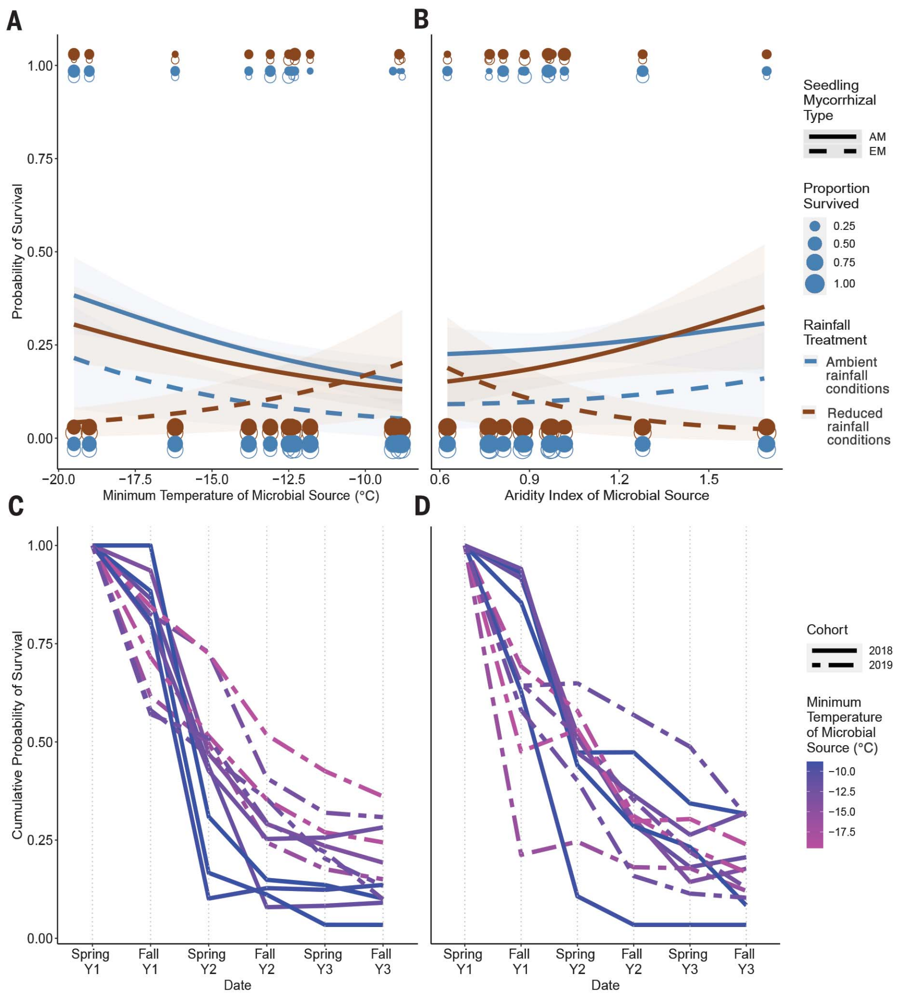


Fig. 2. Seedling survival patterns at the northern field site versus the climatic conditions of the microbial inocula source. (**A** and **B**) Probability of surviving over 3 years versus (A) the minimum temperature or (B) aridity index of the microbial-source site. Aridity index, precipitation/potential evapotranspiration; lower values indicate drier conditions. Blue, ambient rainfall conditions; brown, reduced rainfall conditions. Dashed lines and open symbols, ectomycorrhizal (EM) species; solid lines and filled symbols, arbuscular mycorrhizal (AM) species. (**C** and **D**) Cumulative survival over three years, broken down by spring and fall sampling dates. (C) Ambient rainfall conditions, (D) reduced rainfall conditions. Each line represents the cumulative survival probability for a given microbial inocula source, colored according to the minimum temperature of the microbial-source site. Solid lines, 2018 cohort; dashed lines, 2019 cohort.

Unlike in the northern site, seedlings associated with AM versus EM symbiotic fungi had qualitatively different responses to microbial inoculation sources in our southern site. For AM-associated seedlings, predictions of microbial-induced climate tolerance were strongly supported: AM seedlings survived better when inoculated with microbial communities from hot sources in the ambient, but not reduced, rainfall conditions (PGLMM, P = 0.009 and 0.405, respectively; Fig. 3 and table S6) but showed the opposite pattern with respect to the aridity of the microbial source (PGLMM, P = 0.059 for ambient conditions, P < 0.0001 for rainfall-reduced conditions; Fig. 3 and table S6). Survival of EM-associated seedlings did not display significant relationships in predicted directions with microbial-source climate in any condition (Fig. 3 and table S6).

The impact of microbial-source maximum temperature was broadly distributed across seasonal periods, trending toward a positive relationship in most seasons but not resulting in statistically significant associations with survival for any one seasonal period (PGLMM, P > 0.05 for all; table S7, B and C). By contrast, the effect of microbial-source aridity was concentrated strongly during the second summer. Survival over the second summer depended significantly on the interaction between rainfall treatment and microbial-source aridity (PGLMM, P = 0.014; table S7) because of a strongly significant association in the rainfall-reduction treatment (PGLMM, P = 0.003; table S7) but no significant relationship in the ambient rainfall treatment (PGLMM, P = 0.493; table S7). Thus, preinoculation with microbial communities from more arid sites led to enhanced drought tolerance for seedlings during the period of most intense drought stress (in rainfall-reduction shelters during the summer months).

Microbially mediated tolerance to drought in controlled conditions

Our field experiments could not separate the climatic conditions that seedlings experienced from the other conditions at our sites, such as soil properties. Additionally, the variation in our inoculated microbial communities could have arisen from many differences among source sites. In particular, among our northern set of microbial inocula, colder microbial-source sites also had soil conditions more similar to those of the field site (e.g., higher levels of soil organic matter) compared with those of the warmer source sites. Because microbial communities can be adapted to local soil conditions in ways that promote plant growth (*34*), if our inoculated microbial communities were differentially adapted to the soil properties at our field sites, this could be mistaken for microbially induced climate tolerance. Additionally, microbial communities may vary in their direct effects on seedling health in ways unrelated to climate tolerance, such as through pathogen load. Therefore, we performed pot experiments using sterilized soil from each field site, inoculated with the same microbial communities used in our field experiment and grown at ambient and elevated summer temperatures with sufficient or restricted watering frequency (*33*). We used this experiment to test whether the patterns observed in our field sites were dependent on specific climate conditions or rather represented nonclimatic relationships between microbial communities, abiotic soil conditions, and seedlings.

In controlled conditions, seedling growth in the soils from our northern experimental site was unaffected by the climate of the microbial inoculum source, unlike our field results. This indicates that the survival patterns observed in our field experiment could not be explained by nonclimatic effects of microbial communities, such as variable pathogen loads or differential adaptation of microbial communities to the soils at our northern site (table S8). This experiment could not directly confirm the pattern of microbially mediated cold tolerance because the seedlings in the pot experiment did not experience winter temperatures.

Seedling growth in the soil from our southern site demonstrated patterns similar to those seen in our field experiment. In particular, seedling growth was enhanced in pots inoculated with microbial communities from drier

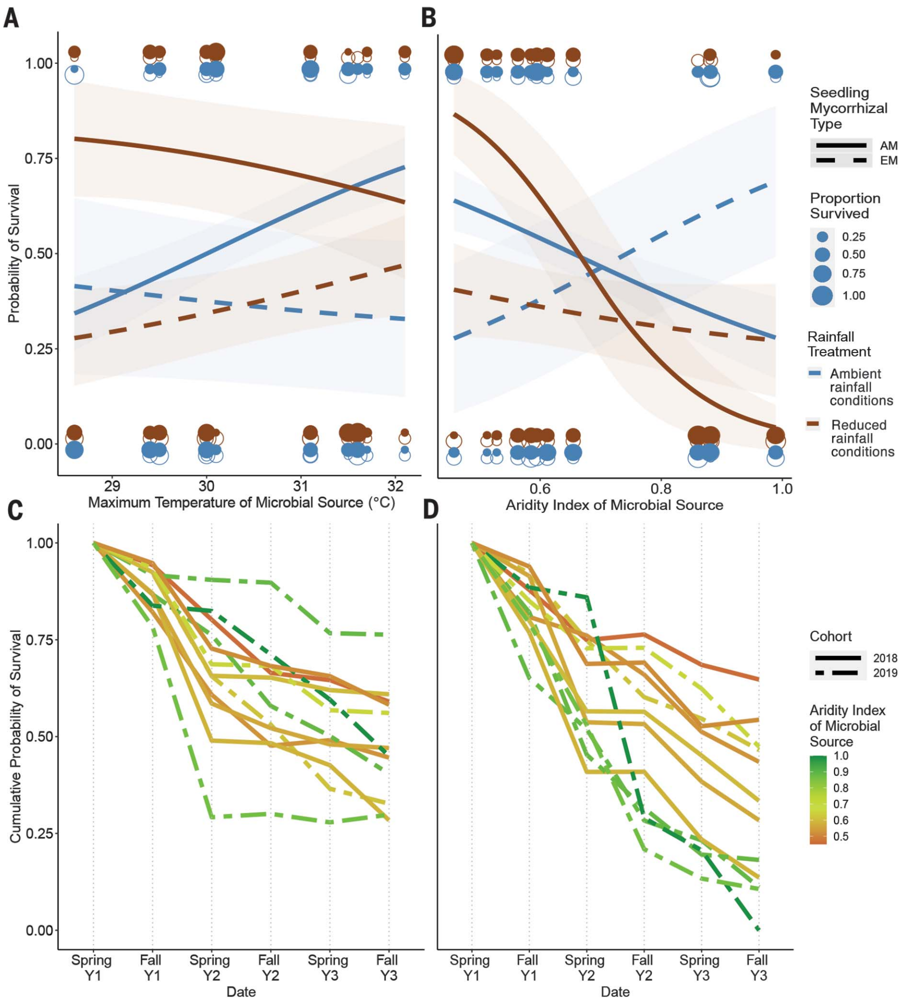


Fig. 3. Seedling survival patterns at the southern field site versus the climatic conditions of the microbial inocula source. Seedling survival at the southern field site. (**A** and **B**) Probability of surviving over 3 years versus the (A) maximum temperature or (B) aridity index of the microbial-source site. Aridity index, precipitation/potential evapotranspiration, measured over the growing season (May to October) prior to soil collection; lower values indicated drier conditions. Blue, ambient rainfall conditions; brown, reduced rainfall conditions. Dashed lines and open symbols, EM species; solid lines and filled symbols, AM species. (**C** and **D**) Cumulative survival over three years, broken down by spring and fall sampling dates. (C) Ambient rainfall conditions, (D) reduced rainfall conditions. Each line represents the cumulative survival probability for a given microbial inocula source, colored according to the aridity index of the microbial-source site. Solid lines, 2018 cohort; dashed lines, 2019 cohort.

sources, but only in the treatment with the greatest drought stress (elevated temperature and restricted watering) (PGLMM, $P < 0.001$; Fig. 4A, table S9, and fig S2). Consistent with the field results, the pattern of microbially mediated drought tolerance was much stronger for AM- compared with EM-associated seedlings (Fig. 4A and table S10).

Microbially mediated climate tolerance is associated with root-associated fungal diversity

Even after 3 years of growth in field conditions, a statistically significant signal of the initial inoculated community remained in the root-associated fungal community composition for both field sites, and for seedlings of both mycorrhizal types [permutational multivariate analysis of variance (perMANOVA), $P < 0.025$ for all; Fig. 5, A and B, figs. S3 to S6, and table S11]. This indicates that at least some fungal taxa from the initial inoculation were able to persist on seedling roots over 3 years, despite the availability of indigenous fungi for replacement, consistent with strong priority effects documented for both AM and EM fungal symbionts (*35*, *36*). Microbially mediated heat and drought tolerance at our southern site appeared to derive directly or indirectly from the diversity of arbuscular mycorrhizal fungi. This inference relies on three observations. First, the diversity of AM fungi was higher in the initial inoculum soil from more arid sites (LM, $P = 0.002$; Fig. 4B and table S12B), and this higher diversity persisted in the roots of surviving AM-associated seedlings even after three years of growth at the field site (PGLMM, $P = 0.043$; Fig. 5C and table S13B). These patterns were not evident in EM-associated seedlings. Second, sources of microbial inocula that resulted in a higher diversity of AM fungi on surviving seedlings also tended to have higher survival probabilities for AM seedlings in the rainfall-reduction treatment, but not in the ambient rainfall treatment (GLM, $P = 0.015$; Fig. 5D and table S14B). Third, these results were confirmed in our controlled-environment experiment, where growth of AM-associated seedlings in the warm-dry treatment was positively correlated to the initial diversity of AM fungi in the soil inoculum (PGLMM, $P = 0.001$; Fig. 4D and table S15), and AM fungal diversity acted as a mediator of the effect of microbial-source aridity on seedling growth (table S15). No other aspects of the full fungal community, or other functional guild subsets, met all three conditions (tables S12B, S13B, and S14B).

Alternatively, enhanced cold tolerance at our northern site occurred regardless of mycorrhizal type. Surviving seedlings inoculated with microbial sources from warmer locations (higher minimum temperature) had a higher richness of fungal species after 3 years at the field site (PGLMM, $P = 0.027$; Fig. 5C and table S14A). In contrast to our southern field site, in the northern site, fungal diversity on surviving seedlings was negatively correlated with the survival probability of seedlings with a given inoculum source in ambient conditions (GLM, $P < 0.001$; Fig. 5C and table S16A). This pattern was evident for richness of the pathogen-endophyte guild but not for either mycorrhizal guild analyzed separately (table S16A). Neither total fungal diversity or richness, nor the diversity or richness of any functional guild, in the original inoculum soil correlated with any source-climate variable (LM, $P > 0.07$ for all; table S13A). Thus, the higher fungal diversity in surviving seedlings inoculated with soils from warmer locations may stem from the combination of inoculated fungal taxa with fungal taxa indigenous to the field site. This suggests that the persistence of maladapted fungal species from warmer-site inocula may have acted to inhibit the provision of cold tolerance by the indigenous fungi.

We expect the precise drivers of microbially induced climate tolerance to be idiosyncratic to particular communities and hosts. However, our main conclusion is that the ability to enhance plant-host tolerance is predictable by the climatic history of the microbial community. Our experiments involved a range of tree species and climate variables, indicating that microbially mediated climate tolerance is

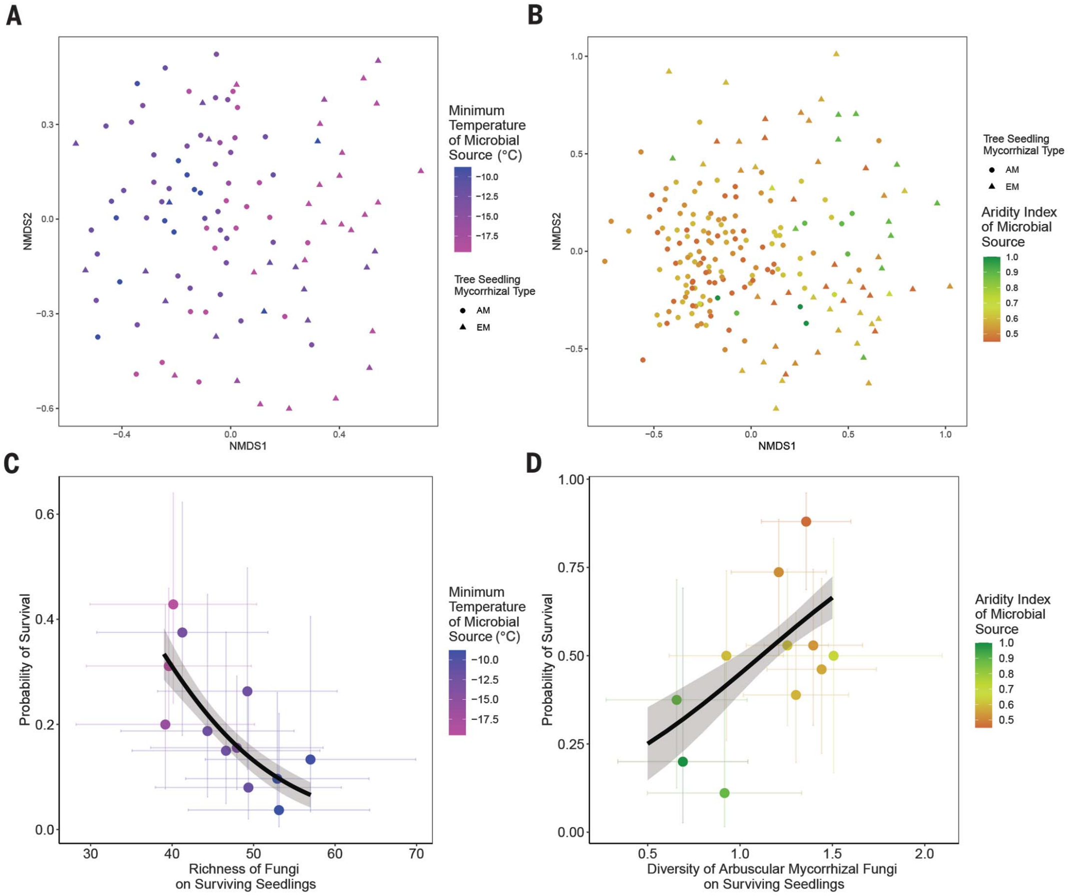


Fig. 4. Fungal communities on surviving seedlings reflect initial microbial inocula sources and correlate with seedling survival probabilities. (**A**) Nonmetric multidimensional scaling (NMDS) ordination of fungal communities on surviving seedlings after three years at the northern field experiment, colored according to the minimum temperature of the initial microbial inocula source. Stress = 0.30. (**B**) NMDS ordination of fungal communities on surviving seedlings at the southern field experiment, colored according to the aridity index of the initial microbial inocula source. Stress = 0.28. (**C**) Probability of survival for seedlings in ambient conditions at the northern field site, per microbial inocula source, versus the average fungal richness of surviving seedlings of that same microbial inocula source. (**D**) Probability of survival for AM seedlings in reduced rainfall conditions at the southern field site, per microbial inocula source, versus the average AM fungal diversity of surviving seedlings of that same microbial inocula source.

not a rare phenomenon in temperate forests. Future research should seek to determine how widely the phenomenon of microbially mediated climate tolerance occurs across ecosystems, plant species, and climatic stresses regardless of the specific microbial mechanism. Additionally, more research is needed to understand how microbial communities will respond to changing climates through internal resorting of species and redistribution of species across landscapes, and what consequence these shifts will have for host plant climate tolerance. For instance, in situ soil warming and drought experiments have documented shifts in microbial community composition (*37–40*), but it is not clear if the resulting new microbial communities will confer increased tolerance to warming or drying soils for plant species. Our experiments demonstrated that over short time periods, introduction of new microbial taxa (through preinoculated tree seedlings) could lead to higher plant survival than the indigenous community could provide on its own for species at their trailing range edge. This suggests that indigenous microbial communities may not be infinitely flexible in terms of functions for host plants and that movement

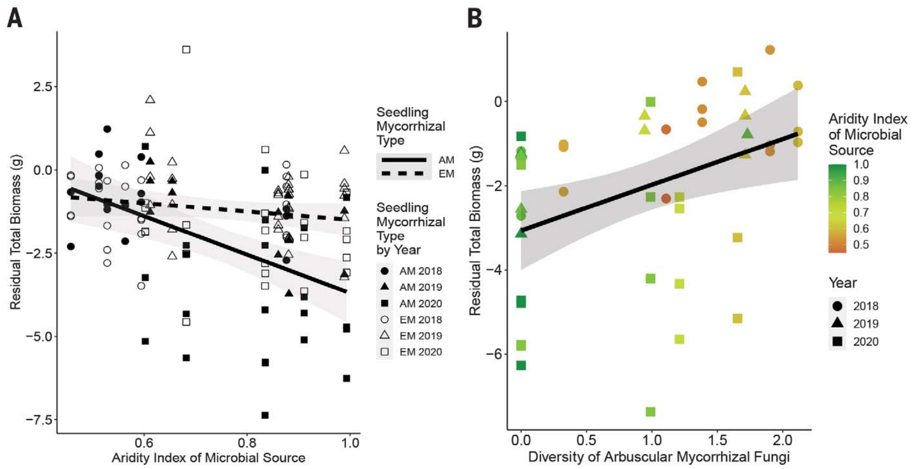


Fig. 5. Seedling growth in water-stressed treatments in controlled conditions reflected microbial-source climate for AM-associated seedlings. Residual seedling biomass in the controlled-condition experiment using southern field-site soil–mirrored patterns seen in the field experiment. Residual values account for differences in growth rate among seedling species and initial size; positive values indicate larger-than-expected biomass for a seedling of that species, whereas negative values indicate a smaller-than-expected biomass. (**A**) Residual biomass of seedlings in the warm-dry treatment versus the aridity index of the initial microbial source, separated by seedling mycorrhizal type. Dashed lines and open symbols, EM species; solid lines and filled symbols, AM species. (**B**) Residual biomass of AM seedlings in the warm-dry condition versus the Shannon-Weaver index of diversity of AM fungi in the original inoculum. Shannon-Weaver diversity of AM fungi in the original inoculum soil was positively correlated with aridity.

of microbial populations across space may be necessary when plants face increasingly stressful conditions. However, the opposite phenomenon may occur at expanding range edges, where conditions are becoming less stressful for a plant species. In this case, introduction of maladapted microbial taxa reduced survival for species in our experiment.

Intentional inoculation of specific microbial taxa into forests is not feasible at scale. However, tree planting is a major component of climate mitigation strategies, with commitments to plant trillions of trees in coming decades (*41*). Trees are typically planted as seedlings after growth in outdoor beds, bringing diverse microbial communities with them to planting sites (*42*). These preinoculated seedlings represent a massive, unintentional movement of rhizosphere taxa across landscapes. However, they may also represent an opportunity for microbially mediated climatic tolerance if tree nurseries were managed to promote colonization by microbial symbionts adapted to specific climatic conditions.

Rapidly changing climates have reshaped ecosystems and driven extinction and will continue to do so. The interconnections in ecosystems, between community members and between communities across space, have the potential to amplify or resist these changes. A better understanding of microbial interactions as an additional strategy for tree populations at expanding and retracting range edges could help inform predictions of forest vulnerability as well as offer new avenues for promoting resilience in managed forests.

REFERENCES AND NOTES

1. C. Román-Palacios, J. J. Wiens, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4211–4217 (2020).
2. M. C. Urban, *Science* **348**, 571–573 (2015).
3. C. D. Thomas *et al*., *Nature* **427**, 145–148 (2004).
4. T. Jezkova, J. J. Wiens, *Proc. Biol. Sci.* **283**, 20162104 (2016).
5. R. Berkelmans, M. J. H. van Oppen, *Proc. Biol. Sci.* **273**, 2305–2312 (2006).
6. R. J. Rodriguez *et al*., *ISME J.* **2**, 404–416 (2008).
7. A. A. H. A. Latef *et al*., *J. Plant Biol.* **59**, 407–426 (2016).
8. S. S. Porter *et al*., *Funct. Ecol.* **34**, 2075–2086 (2020).
9. P. M. Delaux, S. Schornack, *Science* **371**, 796–807 (2021).
10. H. Lambers, C. Mougel, B. Jaillard, P. Hinsinger, *Plant Soil* **321**, 83–115 (2009).
11. I. S. Acuña-Rodríguez, K. K. Newsham, P. E. Gundel, C. Torres-Díaz, M. A. Molina-Montenegro, *Ecol. Lett.* **23**, 1034–1048 (2020).
12. L. M. Márquez, R. S. Redman, R. J. Rodriguez, M. J. Roossinck, *Science* **315**, 513–515 (2007).
13. T. Lehto, J. J. Zwiazek, *Mycorrhiza* **21**, 71–90 (2011).
14. R. Lata, S. Chowdhury, S. K. Gond, J. F. White Jr., *Lett. Appl. Microbiol.* **66**, 268–276 (2018).
15. D. Naylor, D. Coleman-Derr, *Front. Plant Sci.* **8**, 2223 (2018).
16. M. E. Afkhami, P. J. McIntyre, S. Y. Strauss, *Ecol. Lett.* **17**, 1265–1273 (2014).
17. P. D. Stahl, W. K. Smith, *Mycologia* **76**, 261–267 (1984).
18. I. M. Ware *et al*., *Commun. Biol.* **4**, 748 (2021).
19. C. Allsup, R. Lankau, *Ecology* **100**, e02729 (2019).
20. J. A. Lau, J. T. Lennon, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14058–14062 (2012).
21. S. E. Smith, D. J. Read, *Mycorrhizal Symbiosis* (Academic Press, 2008).
22. M. A. Bingham, S. Simard, *Ecosystems* **15**, 188–199 (2012).
23. C. P. terHorst, P. C. Zee, *Funct. Ecol.* **30**, 1062–1072 (2016).
24. S. A. Smith, M. J. Donoghue, *Science* **322**, 86–89 (2008).
25. R. R. Junker, X. He, J. C. Otto, V. Ruiz-Hernández, M. Hanusch, *FEMS Microbiol. Ecol.* **97**, fiab135 (2021).
26. M. J. Choudoir, A. Barberán, H. L. Menninger, R. R. Dunn, N. Fierer, *Ecology* **99**, 322–334 (2018).
27. P. B. Reich *et al*., *Nature* **608**, 540–545 (2022).
28. L. R. Iverson, A. M. Prasad, S. N. Matthews, M. P. Peters, *Ecosystems* **14**, 1005–1020 (2011).
29. R. J. Dial, C. T. Maher, R. E. Hewitt, P. F. Sullivan, *Nature* **608**, 546–551 (2022).
30. C. C. F. Boonman *et al*., *Divers. Distrib.* **28**, 25–37 (2022).
31. G. Forzieri, V. Dakos, N. G. McDowell, A. Ramdane, A. Cescatti, *Nature* **608**, 534–539 (2022).
32. G. B. Bonan, *Science* **320**, 1444–1449 (2008).
33. Materials and methods are available as supplementary materials.
34. N. C. Johnson, G. W. T. Wilson, M. A. Bowker, J. A. Wilson, R. M. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2093–2098 (2010).
35. P. G. Kennedy, K. G. Peay, T. D. Bruns, *Ecology* **90**, 2098–2107 (2009).
36. G. D. A. Werner, E. T. Kiers, *New Phytol.* **205**, 1515–1524 (2015).
37. K. M. DeAngelis *et al*., *Front. Microbiol.* **6**, 104 (2015).
38. M. E. Van Nuland *et al*., *FEMS Microbiol. Ecol.* **96**, fiaa108 (2020).
39. A. Canarini *et al*., *Nat. Commun.* **12**, 5308 (2021).
40. S. D. Frey, R. Drijber, H. Smith, J. Melillo, *Soil Biol. Biochem.* **40**, 2904–2907 (2008).
41. J. Fargione *et al*., *Front. For. Glob. Change* **4**, 629198 (2021).
42. D. Marčiulynienė *et al*., *Microorganisms* **9**, 150 (2021).
43. R. A. Lankau, Data from: Enhanced climate tolerance for trees derived from microbial communities, Dryad (2023);

ACKNOWLEDGMENTS

We thank S. Ewert, V. King, E. Wang, H. Bielski, A. Polman, B. Frick, J. Viere, and K. Griesbaum for research assistance. We thank the staff at the Walnut Street Greenhouse Complex and Kemp Natural Resource Station (University of Wisconsin–Madison), at the Allerton Park and Recreation Center (University of Illinois, Urbana-Champaign), and at the Wisconsin and Illinois Departments of Natural Resources for assistance and access. **Funding:** This work was funded by National Science Foundation award 1651931 to R.A.L. **Author contributions:** R.A.L. conceived of and designed the study, with contributions from C.M.A. and I.G. C.M.A. and I.G. conducted field and laboratory research. R.A.L. analyzed the data. R.A.L. wrote the manuscript with revision by C.M.A. and I.G. I.G. provided artwork. **Competing interests:** The authors declare that they have no competing interests. **Data accessibility:** Raw sequence data for fungal communities are available at the NCBI Short-Read Archive under BioProjects PRJNA944197, PRJNA944945, PRJNA944949, and PRJNA944953. All other data and code are available at the Dryad repository (*43*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2027
Material and Methods
Figs. S1 to S6
Tables S1 to 16
References (*44*–*59*)

View/request a protocol for this paper from *Bio-protocol*.

Submitted 5 October 2022; resubmitted 19 January 2023
Accepted 4 April 2023
10.1126/science.adf2027

THERMOELECTRICS

Lattice plainification advances highly effective SnSe crystalline thermoelectrics

Dongrui Liu[1], Dongyang Wang[2]†, Tao Hong[1]†, Ziyuan Wang[3], Yuping Wang[1], Yongxin Qin[1], Lizhong Su[1], Tianyu Yang[3], Xiang Gao[4], Zhenhua Ge[3], Bingchao Qin[1]*, Li-Dong Zhao[1]*

Thermoelectric technology has been widely used for key areas, including waste-heat recovery and solid-state cooling. We discovered tin selenide (SnSe) crystals with potential power generation and Peltier cooling performance. The extensive off-stoichiometric defects have a larger impact on the transport properties of SnSe, which motivated us to develop a lattice plainification strategy for defects engineering. We demonstrated that Cu can fill Sn vacancies to weaken defects scattering and boost carrier mobility, facilitating a power factor exceeding ~100 microwatts per centimeter per square kelvin and a dimensionless figure of merit (*ZT*) of ~1.5 at 300 kelvin, with an average *ZT* of ~2.2 at 300 to 773 kelvin. We further realized a single-leg efficiency of ~12.2% under a temperature difference (ΔT) of ~300 kelvin and a seven-pair Peltier cooling ΔT_{max} of ~61.2 kelvin at ambient temperature. Our observations are important for practical applications of SnSe crystals in power generation as well as electronic cooling.

The domains of waste-heat recovery, power generation, and electronic refrigeration have substantial application potential for thermoelectric technology, which can enable the direct conversion between heat and electricity (*1*–*5*). The conversion efficiency of thermoelectric technology is determined by the dimensionless figure of merit (*ZT*) of the thermoelectric material, which is defined as $ZT = S^2\sigma T/\kappa_{tot}$ (where σ, S, T, and κ_{tot} represent the electrical conductivity, Seebeck coefficient, absolute temperature, and total thermal conductivity, respectively) (*6*, *7*). A superior average *ZT* (ZT_{ave}) in the operating temperature range is required to achieve high conversion efficiency. Numerous methods have been devised to enhance thermoelectric performance, including optimization of carrier concentration (*8*), manipulation of electronic band structure (*9*–*11*), engineering of full-scale microstructure (*12*–*14*), and decoupling of electron and phonon transports (*15*–*17*).

The simple compound SnSe is one of the most promising thermoelectric candidates with extremely low thermal conductivity owing to its strong bond anharmonicity (*18*, *19*). Subsequently, high-performance p-type and n-type SnSe crystals have been developed, mainly owing to their complex electronic band structure (*20*, *21*) and three-dimensional (3D)–charge and 2D-phonon transport features (*15*, *22*), continuously revealing their potential for thermoelectric applications. Moreover, using the momentum and energy multiband synglisis strategy, the thermoelectric performance, especially near room temperature, was boosted with a power factor (*PF*) of ~75 μW cm^{-1} K^{-2} and a *ZT* of ~1.2 at 300 K in Pb-alloyed p-type SnSe crystals, opening the potential for SnSe crystals to be used in thermoelectric power generation and Peltier cooling (*21*).

To ensure the application of thermoelectric cooling devices, we have greatly reduced the defects concentration and improved the carrier mobility by preparing SnSe crystals (*18*) and modifying the crystal structure (*23*). By manipulating the extrinsic defects (*24*), using multiband synglisis strategy (*21*) and structural modulation (*25*), we further improved the room-temperature thermoelectric performance, but it is still insufficient to be practically used. For SnSe crystalline thermoelectrics, the vast off-stoichiometric defects were revealed in both undoped and doped SnSe crystals, mainly including Sn vacancies (*24*, *26*, *27*). These intrinsic defects impede carrier transport and deteriorate carrier mobility, especially in p-type SnSe crystals (*24*, *27*). Therefore, we intend to manipulate the intrinsic defects (specifically Sn vacancies) in SnSe crystals, which is a feasible way to improve carrier mobility and thermoelectric properties especially at room temperature. Inspired by the concept of "compositional plainification" in the study of structural materials (*28*), we propose a "lattice plainification" strategy, which refers to the use of a proper number of extrinsic atoms to manipulate the intrinsic defects and facilitate the carrier transport, thus realizing high-performance thermoelectrics with a plainer (plainified) lattice. Different from the compositional plainification strategy that realizes intrinsic defect manipulation by simplifying (plaining) chemical compositions (*28*), the lattice plainification strategy can be regarded as its extension in the thermoelectric field. Achieving extraordinary carrier mobility implies higher electrical conductivity, thus contributing partly to lower power dissipations, which is crucial for practical applications of thermoelectric cooling devices. Different from introducing extrinsic defects to improve the carrier concentration and effective mass (*24*), we aim to suppress the defects, especially Sn vacancies, to weaken defects scattering and boost carrier mobility. To fill the Sn vacancies, Cu was selected instead of Sn self-compensation because Cu also can work as hole dopants.

We successfully realized high-performance power generation and Peltier cooling in SnSe crystals using a lattice plainification strategy. By introducing small amounts of Cu upon prior 2% Na doping, we manipulated the intrinsic defects, especially Sn vacancies, favoring a substantial increase in carrier mobility μ (Fig. 1A). Notably, Na doping only generates a high carrier concentration and moves the Fermi level into valence bands, whereas the valence band structure or the crystal lattice structure were not changed (*29*). Specifically, μ was enhanced from ~250 cm^2 V^{-1} s^{-1} to ~319 cm^2 V^{-1} s^{-1} at 300 K compared with that of Cu-free SnSe. In addition to filling Sn vacancies, additional Cu substitutes Sn in the lattice to be p-type doping, increasing the carrier concentration *n* from ~4.0×10^{19} cm^{-3} to ~6.4×10^{19} cm^{-3} at 300 K. Consequently, the electrical conductivity markedly increases from ~1500 S cm^{-1} to ~3047 S cm^{-1} at 300 K, which can be one of the prerequisites for realizing lower thermoelectric device internal resistance and operating power dissipations. Moreover, Cu occupying the Sn positions in the lattice effectively modulates the crystal structure, thus promoting multiband synglisis and further optimizing the effective mass m^* and μ. The synergistically optimized electrical transports lead to the largely enhanced *PF*, exceeding ~100 μW cm^{-1} K^{-2} at 300 K in p-type $SnCu_{0.001}Se$ crystals. Meanwhile, all the samples maintain the intrinsically low thermal conductivities, benefiting high *ZT* values over a broad temperature range. Specifically, an ultrahigh *ZT* of ~1.5 at 300 K with a ZT_{ave} of ~2.2 at 300 to 773 K was achieved, demonstrating the excellent potential for power generation and Peltier cooling. We further designed a single-leg thermoelectric device using $SnCu_{0.001}Se$ samples, and a high conversion efficiency of ~12.2% at a temperature difference ΔT of ~300 K was obtained by using the commercial Mini-PEM instrument. This result outperforms the conversion efficiencies of most reported single-leg and multipair thermoelectric devices under similar temperature differences ΔT (Fig. 1B) (*30*–*36*). Moreover, we fabricated a

[1]School of Materials Science and Engineering, Beihang University, Beijing 100191, China. [2]Henan Key Laboratory of Diamond Optoelectronic Materials and Devices, Key Laboratory of Material Physics, Ministry of Education, School of Physics, Zhengzhou University, Zhengzhou 450052, China. [3]Faculty of Materials Science and Engineering, Kunming University of Science and Technology, Kunming 650093, China. [4]Center for High Pressure Science and Technology Advanced Research (HPSTAR), Beijing 100094, China.
***Corresponding author. Email: qinbingchao@buaa.edu.cn (B.Q.); zhaolidong@buaa.edu.cn (L.-D.Z.)**
†These authors contributed equally to this work.

seven-pair thermoelectric cooling device using p-type $SnCu_{0.001}Se$ crystals coupled with n-type commercial $Bi_2Te_{2.7}Se_{0.3}$. The maximum cooling temperature difference ΔT_{max} reaches ~61.2 K when the hot-end temperature T_h was fixed at ~300 K (Fig. 1C), which is comparable to the cooling performance of commercial Bi_2Te_3-based devices (*37*). Furthermore, larger ΔT_{max} values can be realized at higher T_h conditions, and ΔT_{max} approaches ~90.6 K at T_h of ~343 K (Fig. 1C), showing improvement compared with that of the $Sn_{0.91}Pb_{0.09}Se$-based device (*21*), further promoting the potential applications of SnSe crystals for thermoelectric cooling and precise thermal management of electronic devices.

Electrical transports

We prepared a series of $SnCu_xSe$ crystals using a two-step modified Bridgman method and investigated their transport properties along the *b* axis (fig. S1). After the introduction of Cu, the electrical transport performance was improved (Fig. 2). With increasing Cu content, the carrier concentration shows a slight enhancement, whereas the carrier mobility first increases and then decreases (fig. S2). This process can be attributed to the multiple roles of additional Cu in occupying the Sn vacancies and substituting Sn in the lattice. The former weakens the defects-induced carrier scattering to boost the carrier mobility, whereas the latter becomes acceptor doping to increase the hole carrier concentration. The subsequent decrease in carrier mobility may be caused by the extra Cu becoming interstitial atoms, which strongly scatter carriers when x exceeds 0.001. We have also conducted the scanning transmission electron microscopy (STEM) characterization of the $SnCu_{0.002}Se$ sample to confirm this (fig. S3). The simultaneous enhancement of the carrier concentration and mobility results in higher electrical conductivity, exceeding ~3000 S cm^{-1} at 300 K (Fig. 2A). In addition, higher carrier concentrations push the Fermi level deeper and activate more valence bands to be involved in electrical transport, favoring the maintenance of larger Seebeck coefficients (Fig. 2B). As a result of the synergistic adjustment of the electrical conductivity and Seebeck coefficient, we obtained a maximum *PF* value of ~100 μW cm^{-1} K^{-2} at 300 K by slightly adding 0.001 mol Cu (Fig. 2C), which outperforms those of most p-type thermoelectric materials (fig. S4) (*12*, *20*, *21*, *29*, *38*). The electrical performance can be better reflected by the weighted mobility μ_w (*39*), and the large increase in μ_w (Fig. 2D) suggests that the electrical transport properties are optimized by the synergy between μ and m^* owing to the complex energy band structures. We estimated μ and m^* to better interpret the multiple roles after the introduction of Cu. First, the tiny Cu fills Sn vacancies to plain the lattice and lower the defects concentration, thus diminishing the carrier scattering from lattice defects and producing an excellent μ of ~319 cm^2 V^{-1} s^{-1} at 300 K, which is much higher than the optimized μ values based on multiband synglisis (Fig. 2E) (*21*). Second, the extra Cu substitutes for Sn in the lattice, forming hole doping to increase the carrier concentration n (fig. S2). The addition of Cu increases n and causes the Fermi level to further descend into the valence bands, promoting the multiple valence bands transport. Furthermore, Cu/Sn atomic substitutions have obvious effects on the electronic band structure to promote multiband synglisis. These beneficial effects jointly facilitate an increase in the effective mass m^* to ~1.7 m_e, comparable to that derived from the Pb-doped SnSe crystals (Fig. 2F) (*21*), thus realizing relatively large Seebeck coefficients even at higher n. Such exceptional electrical transport properties in SnSe crystals with a plainer lattice by introducing tiny Cu atoms are expected to facilitate high-performance power generation and Peltier cooling over a wide temperature range.

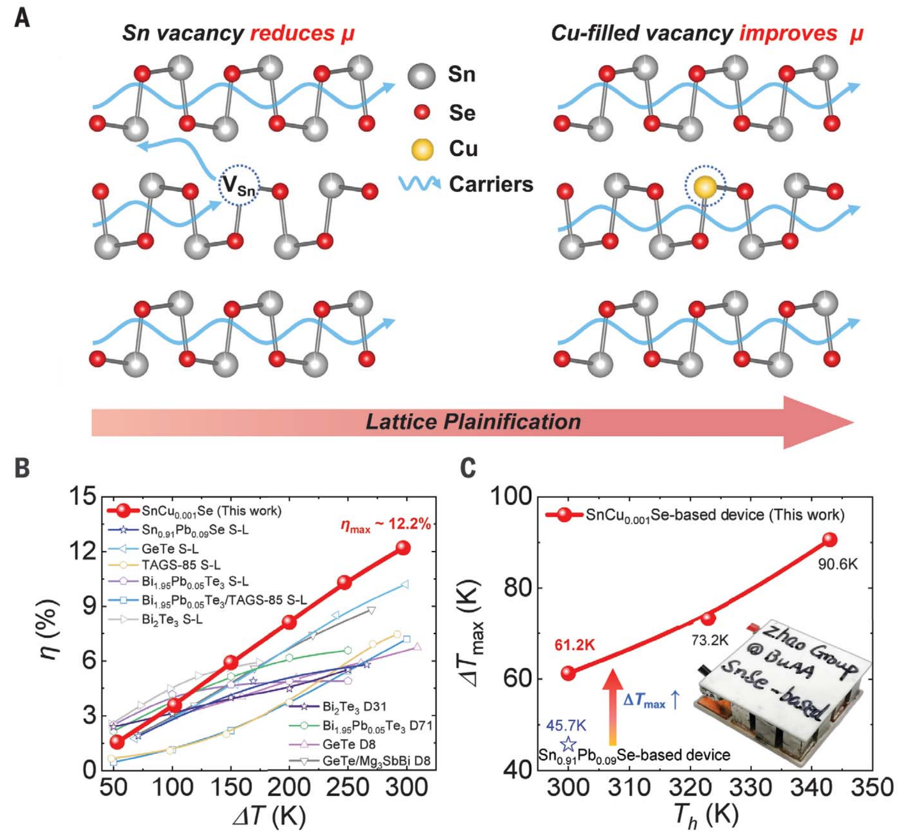


Fig. 1. High-performance power generation and Peltier cooling realized through lattice plainification. (**A**) Intrinsic Sn vacancies in SnSe crystals have negative effects on carrier mobility μ because they strongly scatter carriers. The lattice plainification strategy, by introducing tiny Cu to fill the Sn vacancies and weaken carrier scattering, leads to much higher carrier mobility μ and thermoelectric performance. (**B**) Measured conversion efficiency η of $SnCu_{0.001}Se$-based single-leg device as a function of temperature difference ΔT as well as the reported high-performance conversion efficiencies η of single-leg and multipair devices, including the $(Bi, Sb)_2Te_3$ system (*32*–*35*), the GeTe system (*30*, *31*, *36*), and $Sn_{0.91}Pb_{0.09}Se$ crystals (*21*). (S-L represents single-leg device, and Dx represents x-pair device.) (**C**) Maximum cooling temperature difference ΔT_{max} as a function of the hot-end temperature T_h for the seven-pair thermoelectric cooling device made up of p-type $SnCu_{0.001}Se$ and n-type commercial $Bi_2Te_{2.7}Se_{0.3}$. The ΔT_{max} for the $Sn_{0.91}Pb_{0.09}Se$-based cooling device was also plotted for comparison (*21*). The inset is the full-size $SnCu_{0.001}Se$-based thermoelectric cooling device.

Multiple Cu roles for lattice plainification

We further unraveled the roles of Cu through theoretical and experimental observations. The introduction of Cu initially realizes the lattice plainification effect by modulating intrinsic defects to approach a plainer lattice. We confirmed this through defect-related theoretical calculations from a thermodynamic perspective. Considering the presence of Sn vacancies, we explored various defects, including Cu-occupying Sn sites (Cu_{Sn}), Sn vacancies (V_{Sn}), Cu interstitials (Cu_i), and multiple defects that might

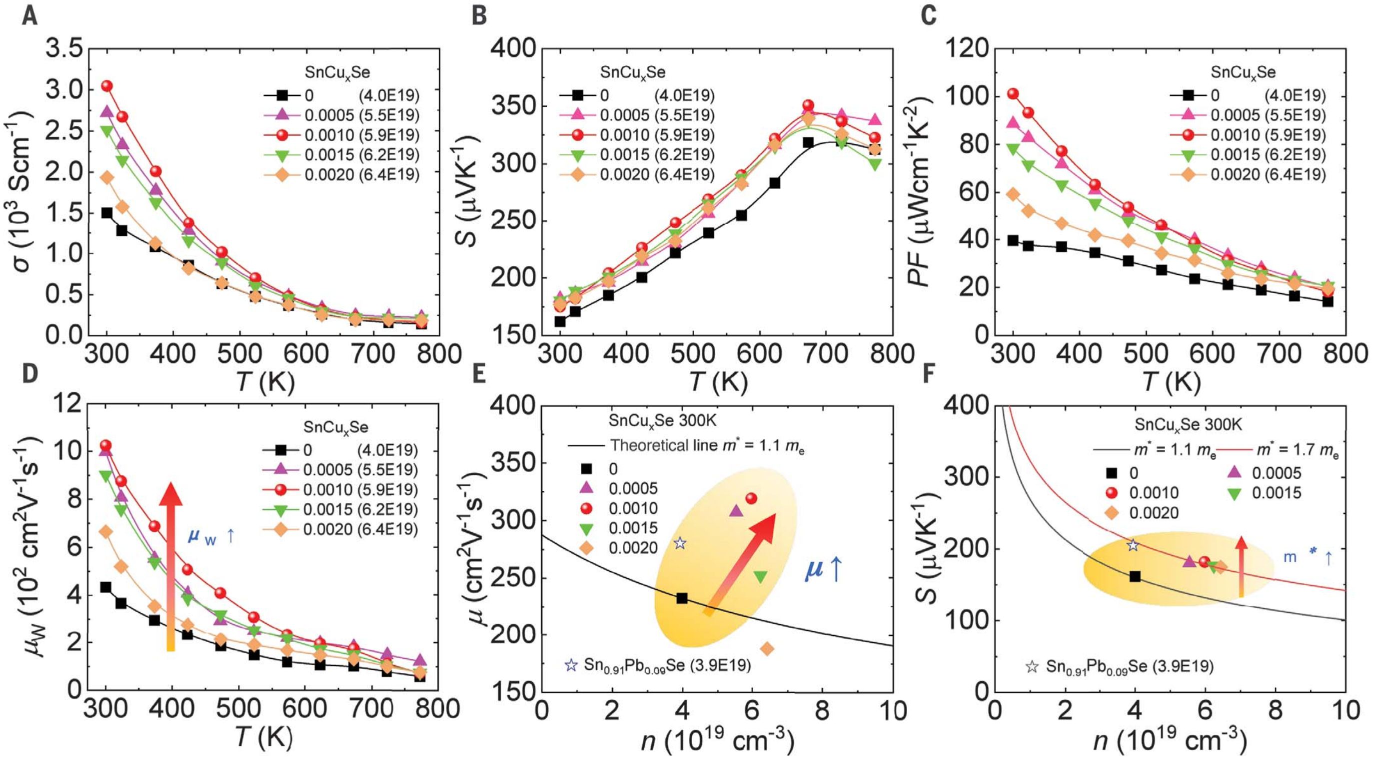


Fig. 2. Electrical transport properties of $SnCu_xSe$. (**A**) Electrical conductivity (σ). (**B**) Seebeck coefficient (*S*). (**C**) Power factor (*PF*). (**D**) Weighted mobility (μ_w). (**E** and **F**) Carrier mobility (μ) (E) and Seebeck coefficient (*S*) (F) as a function of carrier concentration (*n*).

occur after the introduction of Cu. Figure 3, A and B, illustrates the calculated defect formation energy of Cu-related defects under both Sn- and Se-rich conditions. Whether Cu fills Sn vacancies or substitutes Sn, it occupies the Sn site (Cu_{Sn}) in the lattice. The formation energy of Cu_{Sn} is much lower than that of Cu interstitials, Sn vacancies, and other typical defects in both Sn- and Se-rich situations. This suggests that tiny Cu atoms fill the intrinsic Sn vacancies to plain the lattice, which is consistent with the experimental findings that introducing Cu can weaken the carrier scattering from defects and thus achieve an appreciable μ. In addition, as the Cu content increases, extra Cu may readily tend to substitute Sn in the lattice, which slightly improves *n* from ~4.0×10^{19} cm^{-3} to ~6.4×10^{19} cm^{-3}. To gain insight into the kinetic stability of Cu-occupying Sn vacancies, the diffusion energy barrier required for Cu atoms to drift away from the initial interstitial site at the van der Waals gap of SnSe to the neighboring stable Sn vacancy site was determined (Fig. 3C). The energy of a Cu-occupying Sn vacancy is much lower than that of the Cu interstitial and the neighboring Sn vacancy coexisting, and the relatively low energy barrier of ~0.3 eV further verifies the stable state of Cu-occupying Sn vacancies (*40*).

We used the Cs-corrected STEM to directly observe Sn vacancies and Cu atoms at the atomic scale. The contrast of the high-angle annular dark-field (HAADF)–STEM image is proportional to the square of the average atomic number in the sample and has atomic-scale resolution; thus, different atoms can be distinguished efficiently at the atomic scale in close range. Figure S5, A and B, shows high-magnification HAADF-STEM images of Cu-free SnSe along the [100] axis, with a clear weak patch indicating a high density of Sn vacancies. The line-profile intensities (fig. S5C) of the areas marked in fig. S5B indicate the existence of Sn vacancies in Cu-free SnSe. By contrast, fig. S5, D and E, represents the high-magnification HAADF-STEM images of $SnCu_{0.001}Se$ along the same [100] direction. No apparent Sn vacancies were observed in the entire region, but there were some dark contrasts owing to the atomic brightness. The intensity profile of the selected area with minor contrast (fig. S5F) reveals that the Cu atoms occupy the Sn vacancies and have slightly lower intensities compared with Sn atoms. To observe the atomic difference before and after introducing tiny Cu atoms, we illustrated schematics of the atomic site occupation based on the atomic-resolved STEM images (fig. S6). Furthermore, we carried out an elemental analysis of a large area of the $SnCu_{0.001}Se$ sample, and the results are shown in fig. S7, in which the region with enriched Cu can be detected. As indicated by the magnified annular bright-field (ABF)–STEM image in fig. S7D, the atomic arrangement of the Cu-substituted region still maintains the *Pnma* structure of SnSe, which indicates that Cu enters the SnSe lattice to occupy Sn sites.

Moreover, we confirmed that the tiny Cu atoms enter the crystal lattice to occupy the Sn site by investigating the electronic band structures. Based on the refined lattice parameters and atomic positions from the high-temperature synchrotron radiation x-ray diffraction (SR-XRD) experiments (fig. S8 and table S1), we performed density functional theory (DFT) calculations of the electronic band structures of $SnCu_{0.001}Se$ (fig. S9) at various temperatures and obtained the dynamic evolution of the valence bands for the SnSe and $SnCu_{0.001}Se$ (fig. S10) samples. The introduction of Cu substantially affects the band structure mainly by varying the energy differences between the valence band maxima (VBM), and a direct comparison at 300 K is depicted in fig. S11. Tiny Cu atoms occupying Sn slightly alter the valence band structure (Fig. 3D) by decreasing the energy gaps between the first VBM (VBM 1) and other VBMs (VBMs 2, 3, and 4). After we experimentally investigated and verified the effects on electronic band structures after introducing Cu through designing the Cu-control SR-XRD experiments, we obtained a general trend of changes in electronic band structures to

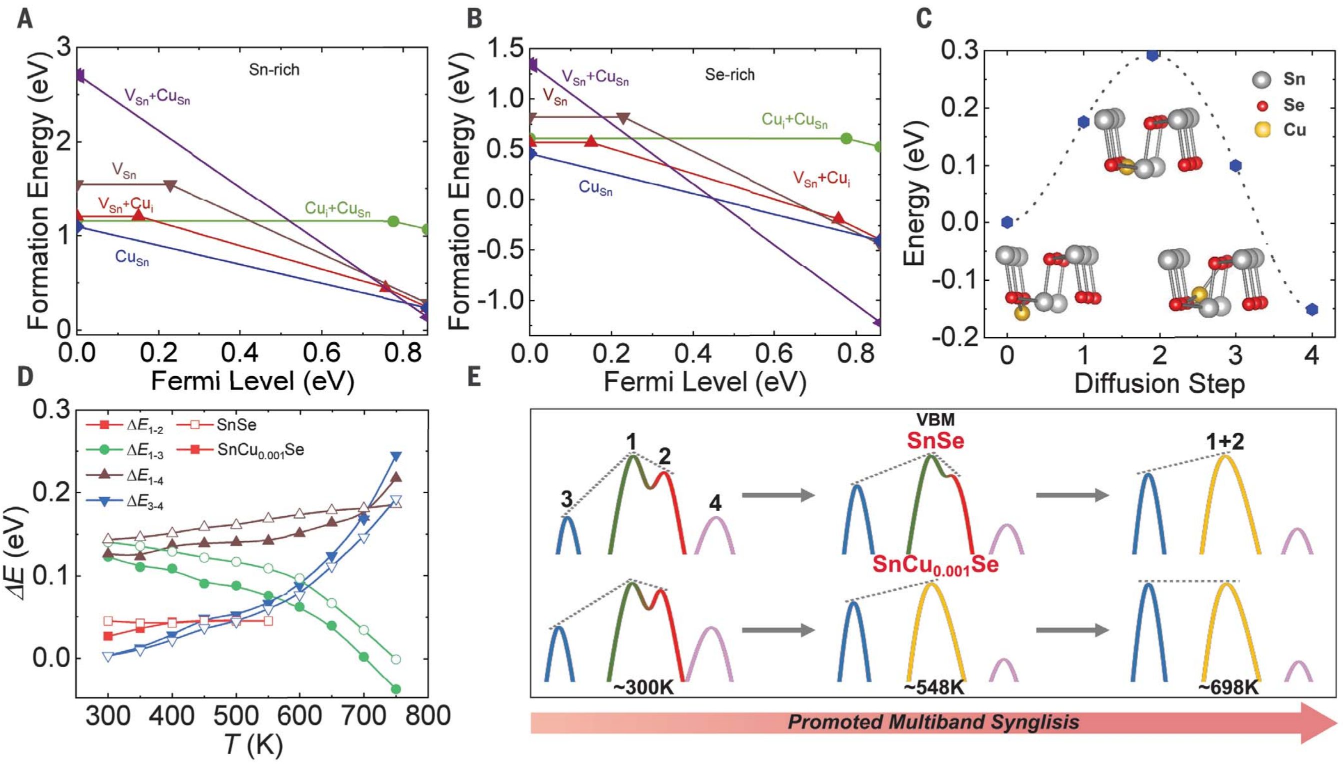


Fig. 3. Multiple Cu roles revealed by theoretical simulations on the defect formation energy and valence band structure in $SnCu_{0.001}Se$. (**A** and **B**) The calculated defect formation energy for SnSe after introducing Cu under Sn-rich (A) and Se-rich (B) conditions considering the presence of Sn vacancies. (**C**) Minimum energy pathway for the diffusion of the introduced Cu atom from the interstitial site to the stable neighboring Sn vacancy site. The initial state refers to the interstitial Cu atom in the van der Waals interlayer with a neighboring Sn vacancy, the intermediate state refers to the Cu position with highest energy, and the final state represents Cu occupying the neighboring Sn vacancy. (**D**) Energy differences (ΔE) between the four VBMs as a function of temperature for p-type SnSe and $SnCu_{0.001}Se$. (**E**) Schematic of the valence bands evolution with temperature for SnSe and $SnCu_{0.001}Se$ demonstrates that introducing Cu slightly promotes the multiband synglisis in p-type SnSe. This promotion was reflected by estimating the typical temperatures during the multiband evolution. The introduction of Cu lowers the momentum alignment temperature of VBM 1 and VBM 2 merging into VBM (1+2) from ~573 K to ~548 K and lowers the energy alignment temperature of VBM (1+2) aligning with VBM 3 from ~723 K to ~698 K.

support the above experimental conclusions by constructing SnSe supercells with Sn vacancies (Sn_{35} Se_{36}) and Cu-occupying Sn vacancies ($Sn_{35}CuSe_{36}$) (fig. S12, A and B). We observed that Cu filling Sn vacancies slightly diminished the energy difference between the valence maxima along the Γ-Z and Γ-Y directions, consistent with the electronic band structure differences calculated from the SR-XRD experimental data, further confirming that Cu fills Sn vacancies and occupies Sn sites in the lattice. The atomic contributions to the obtained band structures also verified that Cu filling Sn vacancies mainly contributes to the valence bands (fig. S12, C and D). The microstructural observations, SR-XRD experiments, and DFT calculations all demonstrate that tiny Cu atoms contribute to the realization of lattice plainification and boost carrier mobility by filling Sn vacancies in the lattice.

In addition to modulating the defects to realize lattice plainification, we observed that the introduction of Cu tunes the band structure and slightly promotes multiband synglisis in p-type SnSe (*21*), which substantially optimizes μ and m^*. First, considering the four VBMs of SnSe, the energy difference between the VBM 1 and VBM 2 along the Γ-Z direction is only ~0.06 eV (Fig. 3D), whereas the energy differences between VBM 1 and VBMs 3 and 4 along the Γ-Y and U-X directions are also relatively low, at ~0.13 eV. These small energy differences make it easy to activate multiple valence bands to participate in carrier transport, thus enlarging m^* and the Seebeck coefficient. Second, the valence band evolution of multiband synglisis contains the bands merging from VBM 1 and VBM 2 to become VBM (1+2) and band alignment between VBM 3 and VBM (1+2), which has been thoroughly investigated in our previous study (*21*). Additionally, we discovered that, compared with Cu-free SnSe, the introduction of such tiny Cu accelerates the dynamic evolution of the multiband synglisis by lowering the bands merging temperature of VBM 1 and VBM 2 becoming VBM (1+2) from ~573 K to ~548 K and the bands alignment temperature between VBM (1+2) and VBM 3 from ~723 K to ~698 K (Fig. 3E). The upgraded valence band evolution can be attributed to the modified crystal structure from the low-symmetry *Pnma* phase to approaching the high-symmetry *Cmcm* phase, as reflected by the chosen angle in the crystal structure (fig. S13). The modified crystal structure might be further due to Cu substituting Sn in the lattice because CuSe stabilizes in a high-symmetry *Cmcm* phase (*41*, *42*). The promoted multiband synglisis along with the lattice plainification by introducing tiny amounts of Cu contribute substantially to optimizing the carrier mobility and effective mass, providing much higher *PF* values over a wide temperature range (Fig. 2C).

Thermal transports, *ZT*, and device efficiency

In addition to the enhancement in electrical transport properties, the thermal conductivity remained relatively low over the entire temperature range (Fig. 4A). The total thermal conductivity κ_{tot} shows a slight increase after

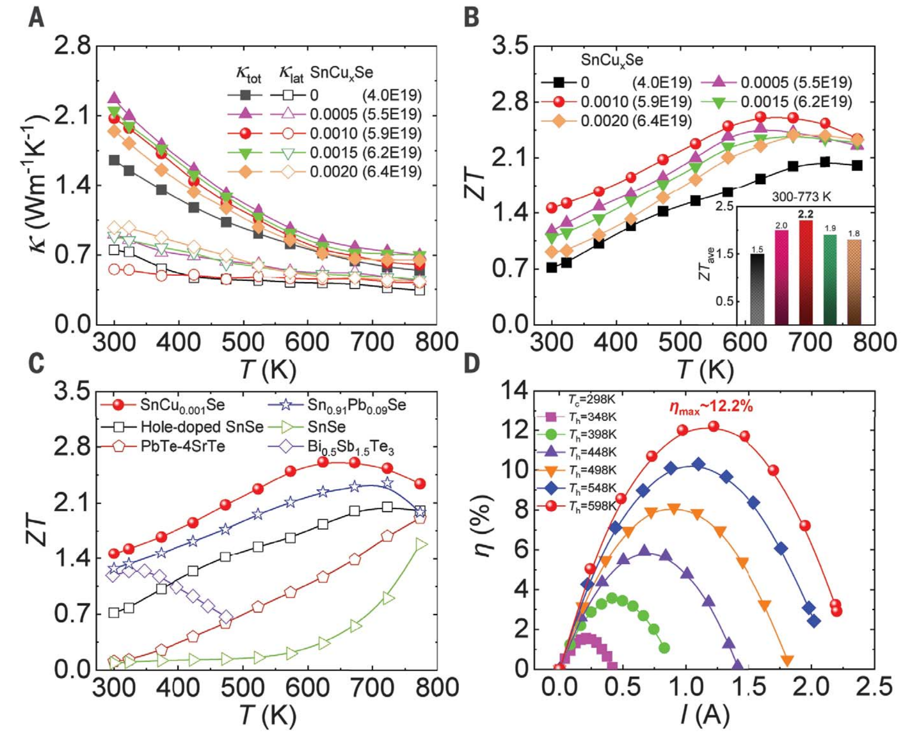


Fig. 4. Thermal transports, dimensionless figure of merit *ZT*, and power generation. (**A**) Total thermal conductivity (κ_{tot}) and lattice thermal conductivity (κ_{lat}) of $SnCu_xSe$ crystals as a function of temperature. (**B**) *ZT* values of $SnCu_xSe$ crystals as a function of temperature, with the inset showing the average *ZT* values (ZT_{ave}) at 300 to 773 K. (**C**) Comparison of *ZT* curves between the $SnCu_{0.001}Se$ crystal in this study and other high-performance p-type thermoelectrics, including $Sn_{0.91}Pb_{0.09}Se$ (*21*), hole-doped SnSe (*20*), undoped SnSe (*18*), PbTe-4SrTe (*12*), and $Bi_{0.5}Sb_{1.5}Te_3$ (*38*). (**D**) Measured conversion efficiency η for $SnCu_{0.001}Se$-based single-leg device as functions of current *I* at different hot-end temperatures T_h, with the cold-end temperature T_c fixed at ~298 K.

adding tiny Cu atoms, and the sample with 0.0005 mol Cu shows a maximum κ_{tot} value of ~2.3 W m^{-1} K^{-1} at 300 K (Fig. 4A). κ_{tot} can be divided into the electronic (κ_{ele}) and lattice (κ_{lat}) parts ($\kappa_{tot} = \kappa_{ele} + \kappa_{lat}$, and $\kappa_{ele} = L\sigma T$, where *L* represents the Lorenz number), and the related parameters to determine thermal transport, including the thermal diffusivity *D*, specific heat C_p, *L*, and κ_{ele}, are depicted in fig. S14. The increase in κ_{tot} originates from the much higher κ_{ele} value, which substantially corresponds to enhanced electrical conductivity (Fig. 2A). Moreover, the κ_{lat} value remains extremely low because of the intrinsic strong anharmonicity in SnSe, and κ_{lat} values at 300 K for all samples locate lower than 1.0 W m^{-1} K^{-1}, with the minimum κ_{lat} value even lower at ~0.6 W m^{-1} K^{-1} for the sample with 0.001 mol Cu, comparable to that of the polycrystalline SnSe after reagent purification and tin oxide removal (*43*).

The lattice plainification strategy of introducing tiny Cu into SnSe essentially realizes the synergy between electrical and thermal transports, leading to ultrahigh thermoelectric performance and *ZT* values. Specifically, in the sample with 0.001 mol Cu, the maximum *ZT* value of ~1.5 was attained at 300 K (Fig. 4B), with a record-high average *ZT* (ZT_{ave}) of ~2.2 at 300 to 773 K (Fig. 4B, inset). The obtained *ZT* curves exceed most of the state-of-the-art thermoelectrics (Fig. 4C) (*12*, *18*, *20*, *21*, *29*, *38*). Notably, in addition to thermoelectric performance optimization, the strategy to plain the lattice by adding tiny Cu boosts the mechanical hardness and machinability of SnSe crystals, and the microhardness after lattice plainification (~1.72 GPa) considerably exceeds those of Bi_2Te_3 ingots (~0.83 GPa) and of undoped (~1.45 GPa) and hole-doped SnSe crystals (~1.52 GPa) (fig. S15). Our high-performance $SnCu_{0.001}Se$ crystals also demonstrated exceptional repeatability and thermal stability (figs. S16 and S17).

The exceptional thermoelectric performance and advantageous machinability further motivated us to fabricate high-performance power generation and thermoelectric cooling devices. During device assembly, the interface between the metallization layer and thermoelectric component, also known as the electrode contact interface, is currently the major bottleneck that limits device performance. Especially for crystalline thermoelectrics, the development of crystal-based thermoelectric devices has been hampered because of the high contact resistance and low bonding strength of the interface, making it essential to improve device performance by interfacial design. Following our previous study using Ni as the metallization layer (*21*), we used two methods to optimize contact resistance. First, a Sn layer was preplated between the SnSe crystal and Ni layer to increase the interfacial wettability (*44*). Second, an Au layer was sandwiched between the Ni layer and solder to optimize the interfacial bonding strength and contact resistance (*45*). The experimental details of preparing contact layers for handling the upper and lower surfaces of the samples can be found in the supplementary materials. Guided by these approaches, we prepared a single-leg thermoelectric device with p-type $SnCu_{0.001}Se$ and measured its power-generation performance using a commercial Mini-PEM instrument (fig. S18). A superior energy conversion efficiency η of ~12.2% under a temperature difference ΔT of ~300 K was achieved (Fig. 4D), which is almost the highest value among the reported high-performance thermoelectric devices at similar temperature differences (Fig. 1B), and an output power of ~166 mW was realized (fig. S19). The theoretical conversion efficiency is calculated to be ~16.2% at the same ΔT (fig. S20A)—the discrepancy between the experimental and theoretical efficiencies mainly comes from the instability of the interfacial contacting materials, the large contact resistance, and the inevitable heat loss due to the size mismatch between the sample and heating stage during the measurement. Furthermore, we constructed a seven-pair thermoelectric cooling device using $SnCu_{0.001}Se$ crystals as the p-type legs and commercial $Bi_2Te_{2.7}Se_{0.3}$ as the n-type legs (thermoelectric performance depicted in fig. S21) and then used the commercial Z-Meters equipment to estimate the thermoelectric cooling performance of the device. Consequently, our device produced a maximum cooling temperature difference ΔT_{max} ~ 61.2 K when the hot-end temperature T_h was fixed at ~300 K (Fig. 1C and fig. S22), demonstrating a cooling performance comparable to that of Bi_2Te_3-based thermoelectric coolers and superior to those of Mg-based alloys (*37*, *46*–*48*). Further, higher ΔT_{max} values of ~73.2 K and ~90.6 K can be achieved at T_h of ~323 K (fig. S23) and ~343 K (fig. S24), respectively, showing considerable application advantages at temperatures above 300 K. At ambient temperature, the experimental ΔT_{max} has reached 71.7% of the theoretical ΔT_{max} (fig. S20B). We believe that the gap between theoretical and experimental ΔT_{max} values can be further narrowed by optimizing the interfacial contacting structure and the size of thermoelectric particles in the device (*49*, *50*). Therefore, much higher

performance in both thermoelectric power generation and Peltier cooling are expected through future optimizing of the contact resistance and shape factor, which are believed to be the critical factors that determine the performance of thermoelectric devices. Additionally, compared with the commercial Bi_2Te_3-based alloys, SnSe crystals have lower costs, more abundant elements, and better mechanical hardness and processability—all of which facilitate the further applications of SnSe crystals with a plainer lattice as thermoelectric coolers.

Conclusion

We developed a lattice plainification strategy to manipulate the intrinsic defects in p-type SnSe crystals. The introduced tiny Cu atoms fill the Sn vacancies to plain the lattice, substitute Sn to generate more hole carriers, and modify the crystal structure to promote multiband synglisis, facilitating exceptional *PF* and *ZT* values, especially at low temperatures. Accompanied by the interfacial design of the thermoelectric devices, we realized high-performance power generation and Peltier cooling in the $SnCu_{0.001}Se$-based devices. Additionally, the improved hardness and machinability imply the superiority of the lattice plainification strategy in practical applications. Our study suggests that we can design better devices based on high-performance $SnCu_xSe$ crystals and explore more high-performance systems through the lattice plainification strategy.

REFERENCES AND NOTES

1. Q. Yan, M. G. Kanatzidis, *Nat. Mater.* **21**, 503–513 (2022).
2. B. Qin, L. D. Zhao, *Science* **378**, 832–833 (2022).
3. J. Mao, G. Chen, Z. Ren, *Nat. Mater.* **20**, 454–461 (2021).
4. X. L. Shi, J. Zou, Z. G. Chen, *Chem. Rev.* **120**, 7399–7515 (2020).
5. T. Zhu *et al.*, *Adv. Mater.* **29**, 1605884 (2017).
6. G. Tan, L. D. Zhao, M. G. Kanatzidis, *Chem. Rev.* **116**, 12123–12149 (2016).
7. X. Tang, Z. Li, W. Liu, Q. Zhang, C. Uher, *Interdiscip. Mater.* **1**, 88–115 (2022).
8. Y. Xiao, L. D. Zhao, *NPJ Quantum Mater.* **3**, 55 (2018).
9. Y. Pei *et al.*, *Nature* **473**, 66–69 (2011).
10. J. P. Heremans *et al.*, *Science* **321**, 554–557 (2008).
11. W. Liu *et al.*, *Phys. Rev. Lett.* **108**, 166601 (2012).
12. K. Biswas *et al.*, *Nature* **489**, 414–418 (2012).
13. B. Jiang *et al.*, *Science* **371**, 830–834 (2021).
14. Y. Jiang *et al.*, *Nat. Commun.* **13**, 6087 (2022).
15. L. Su *et al.*, *Science* **375**, 1385–1389 (2022).
16. S. Roychowdhury *et al.*, *Science* **371**, 722–727 (2021).
17. W. Zhao *et al.*, *Nature* **549**, 247–251 (2017).
18. L. D. Zhao *et al.*, *Nature* **508**, 373–377 (2014).
19. C. W. Li *et al.*, *Nat. Phys.* **11**, 1063–1069 (2015).
20. L. D. Zhao *et al.*, *Science* **351**, 141–144 (2016).
21. B. Qin *et al.*, *Science* **373**, 556–561 (2021).
22. C. Chang *et al.*, *Science* **360**, 778–783 (2018).
23. B. Qin *et al.*, *J. Am. Chem. Soc.* **141**, 1141–1149 (2019).
24. B. Qin *et al.*, *J. Am. Chem. Soc.* **142**, 5901–5909 (2020).
25. B. Qin *et al.*, *Nat. Commun.* **14**, 1366 (2023).
26. D. Wu *et al.*, *Nano Energy* **35**, 321–330 (2017).
27. Z. Wang *et al.*, *Nat. Commun.* **9**, 47 (2018).
28. X. Li, K. Lu, *Science* **364**, 733–734 (2019).
29. W. He *et al.*, *Science* **365**, 1418–1424 (2019).
30. B. Jiang *et al.*, *Science* **377**, 208–213 (2022).
31. Z. Bu *et al.*, *Sci. Adv.* **7**, eabf2738 (2021).
32. A. Muto, D. Kraemer, Q. Hao, Z. F. Ren, G. Chen, *Rev. Sci. Instrum.* **80**, 093901 (2009).
33. F. Hao *et al.*, *Energy Environ. Sci.* **9**, 3120–3127 (2016).
34. A. K. Bohra *et al.*, *Mater. Des.* **159**, 127–137 (2018).
35. B. Zhu *et al.*, *Energy Environ. Sci.* **13**, 2106–2114 (2020).
36. Z. Bu *et al.*, *Energy Environ. Sci.* **14**, 6506–6513 (2021).
37. J. Yang *et al.*, *Joule* **6**, 193–204 (2022).
38. R. Deng *et al.*, *Sci. Adv.* **4**, eaar5606 (2018).
39. G. J. Snyder *et al.*, *Adv. Mater.* **32**, 2001537 (2020).
40. D. Young, *Computational Chemistry: A Practical Guide For Applying Techniques to Real World Problems* (Wiley, 2004).
41. V. Milman, *Acta Cryst.* **B58**, 437–447 (2002).
42. R. D. Heyding, *Can. J. Chem.* **44**, 1233–1236 (1966).
43. C. Zhou *et al.*, *Nat. Mater.* **20**, 1378–1384 (2021).
44. T.-H. Chuang, W.-T. Yeh, C.-H. Chuang, J.-D. Hwang, *J. Alloys Compd.* **613**, 46–54 (2014).
45. Y. N. Nguyen, K.-T. Kim, S.-H. Chung, I. Son, *J. Alloys Compd.* **895**, 162716 (2022).
46. B. Poudel *et al.*, *Science* **320**, 634–638 (2008).
47. J. Mao *et al.*, *Science* **365**, 495–498 (2019).
48. Z. Liu *et al.*, *Nat. Commun.* **13**, 1120 (2022).
49. H. Wang, *Mater. Lab* **2**, 220053 (2022).
50. Y. Qin, B. Qin, D. Wang, C. Chang, L.-D. Zhao, *Energy Environ. Sci.* **15**, 4527–4541 (2022).

ACKNOWLEDGMENTS

The authors thank BL14B1 (Shanghai Synchrotron Radiation Facility) for the SR-XRD experiments. **Funding:** This work was supported by the Basic Science Center Project of the National Natural Science Foundation of China (52388201), the National Natural Science Foundation of China (52250090), the National Key Research and Development Program of China (2018YFA0702100), the National Natural Science Foundation of China (51571007, 51772012, and 12204156), the Beijing Natural Science Foundation (JQ18004), and the 111 Project (B17002). L.-D.Z. acknowledges support from the National Science Fund for Distinguished Young Scholars (51925101) and the high-performance computing (HPC) resources at Beihang University. B.Q. acknowledges support from the Academic Excellence Foundation of BUAA for PhD students. **Author contributions:** D.L., B.Q., and L.-D.Z. synthesized the samples, designed, and carried out the experiments, analyzed the results, and wrote the paper. D.L., Y.W., and L.S. carried out the high-temperature SR-XRD experiments and Rietveld refinements. D.W. carried out the DFT calculations. T.H. and X.G. conducted the STEM experiments. D.L., Z.W., and Z.G. performed the single-leg device fabrication and measured the power generation performance. T.Y. and Z.G. measured the Vickers hardness. D.L., B.Q., and Y.Q. built the seven-pair thermoelectric cooling device and performed the cooling performance measurements. All authors conceived the experiments, analyzed the results, and coedited the manuscript. **Competing interests:** D.L., B.Q., and L.-D.Z. are inventors on a patent (202310134387.9) submitted by Beihang University that covers the topic of the synthesis technology for the p-type SnSe crystals and the use for new thermoelectric cooling candidates. The authors declare no other competing interests. **Data and materials availability:** All data are provided in the main text or the supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg7196
Materials and Methods
Figs. S1 to S24
Tables S1 and S2
References (*51*–*56*)

Submitted 30 January 2023; accepted 7 April 2023
10.1126/science.adg7196

HIBERNATION

Climate change is altering the physiology and phenology of an arctic hibernator

Helen E. Chmura[1,2]*, Cassandra Duncan[3], Grace Burrell[3], Brian M. Barnes[1], C. Loren Buck[4], Cory T. Williams[5]*

Climate warming is rapid in the Arctic, yet impacts to biological systems are unclear because few long-term studies linking biophysiological processes with environmental conditions exist for this data-poor region. In our study spanning 25 years in the Alaskan Arctic, we demonstrate that climate change is affecting the timing of freeze-thaw cycles in the active layer of permafrost soils and altering the physiology of arctic ground squirrels (*Urocitellus parryii*). Soil freeze has been delayed and, in response, arctic ground squirrels have delayed when they up-regulate heat production during torpor to prevent freezing. Further, the termination of hibernation in spring has advanced 4 days per decade in females but not males. Continued warming and phenological shifts will alter hibernation energetics, change the seasonal availability of this important prey species, and potentially disrupt intraspecific interactions.

Climate change is particularly rapid in the Arctic (*1*), where systematic warming is reducing sea ice extent, altering hydrological cycles, thawing permafrost, increasing shrubs, and changing the timing of key seasonal events (phenological shifts) (*2*). Despite the rapid pace of climate change in the Arctic, it is a relatively data-poor region (*3*), and few long-term records combining physical records of climate change and physiological responses of organisms exist [but see (*4*, *5*)]. Further, although changes in the spring and summer have received considerable attention, recent work has called for more research into the consequences of warmer and wetter winters (*6*). Winter conditions shape life histories, because many animal species have evolved strategies such as seasonal migration or dormancy to cope with prolonged periods of low food availability (*7*). In resident species that hibernate, climate change could lead to changes in energy expenditure and overwinter survival. Energy requirements increase markedly when hibernacula temperatures drop below an animal's thermal set point [near freezing for ground squirrels (*8*)], because animals must produce heat to prevent tissue damage and death (*9*). Winter temperatures may also contribute to the regulation of spring life history events (*10*, *11*), and phenological shifts can have important ecological repercussions if they result in mismatches such that historically synchronous interactions within or among species are no longer temporally aligned (*12*, *13*).

[1]Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. [2]Rocky Mountain Research Station, United States Forest Service, Missoula, MT 59801, USA. [3]Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. [4]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA. [5]Department of Biology, Colorado State University, Fort Collins, CO 80523, USA.
*Corresponding author. Email: cory.williams@colostate.edu (C.T.W.); helen.chmura@usda.gov (H.E.C.)

As part of a 25-year field study of arctic ground squirrels (*Urocitellus parryii*) in the Alaskan Arctic, we evaluated the impacts of climate change on this high-latitude mammalian hibernator, focusing on changes in hibernation physiology and emergence phenology. Unlike hibernators in most temperate or montane regions, which sequester themselves in hibernacula that remain above freezing, arctic ground squirrels overwintering in frozen soils must defend themselves against large thermal gradients (or differences between ambient and body temperatures) while torpid using nonshivering thermogenesis (thermogenic torpor) (*9*, *14*) (Fig. 1). We demonstrate that significant warming in ambient air and soil (hibernacula) temperatures is altering hibernation phenology and the duration of thermogenic torpor in this arctic species.

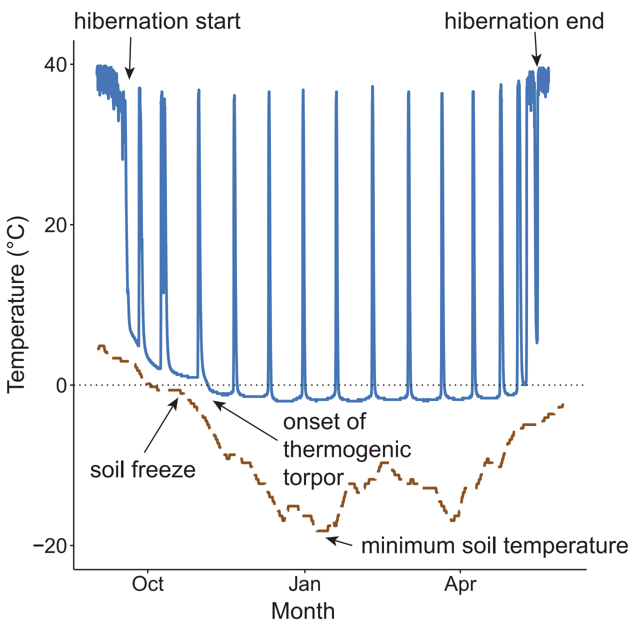


Fig. 1. Seasonal changes in arctic ground squirrel body temperature and soil temperature of the hibernaculum. Abdominal body temperature of a hibernating arctic ground squirrel (blue line) and temperature of the surrounding hibernaculum (soil temperature at 1-m depth adjacent to the burrow entrance, dashed brown line) from a site near Toolik Field Station. Squirrels expend energy to maintain their body temperature above that of the hibernaculum ("thermogenic torpor") such that their brain temperatures never drop below 0°C (*9*).

Climate-driven changes in soil conditions alter hibernation

We used long-term climate records of air and soil temperature from two field sites, Toolik (68°38′ N, 149°38′ W; elevation 719 m) and Atigun (68°27′ N, 149°21′ W; elevation 812 m), in the Alaskan Arctic to document recent environmental change. The dataset durations were as follows: air temperatures at Toolik were measured from fall 1993 to spring 2020, Toolik soil temperatures from fall 1993 to spring 2019, and Atigun soil temperatures from fall 2002 to spring 2019. These measures were paired with hibernation records collected using biologgers (Toolik, fall 1996 to spring 2019; Atigun, fall 1999 to spring 2021) (*15*) to evaluate the physiological impact of recent climate change on arctic ground squirrels (see the supplementary materials). The use of biologgers that measured abdominal and/or skin temperature allowed us to record detailed information about physiological and phenological events during hibernation (Fig. 1, fig. S1, and table S1) from 199 free-living arctic ground squirrel individuals over 25 years.

Average annual air temperatures increased from 1994 to 2020 ($\beta = 0.070$, $t = 2.552$, $P = 0.02$; Fig. 2A). This was driven by increases in winter temperatures, because seasonal analyses revealed increases in winter temperatures ($\beta = 0.119$, $t = 2.375$, $P = 0.025$; fig. S2), but no annual trends for spring, summer, or fall (see the supplementary materials and fig. S2). Dates of soil freeze, measured at 1-m depth adjacent to hibernacula, shifted to be ~4 days/decade later in the fall ($n = 446$, $\beta = 0.409$, $t = 2.851$, df = 383, $P = 0.0045$; Fig. 2B), and minimum soil temperatures in winter increased by almost 2°C/decade ($n = 366$, $\beta = 0.185$, $t = 6.922$, $P < 0.001$; Fig. 2C). Further, dates of soil thaw in summer advanced by ~0.3 days/decade

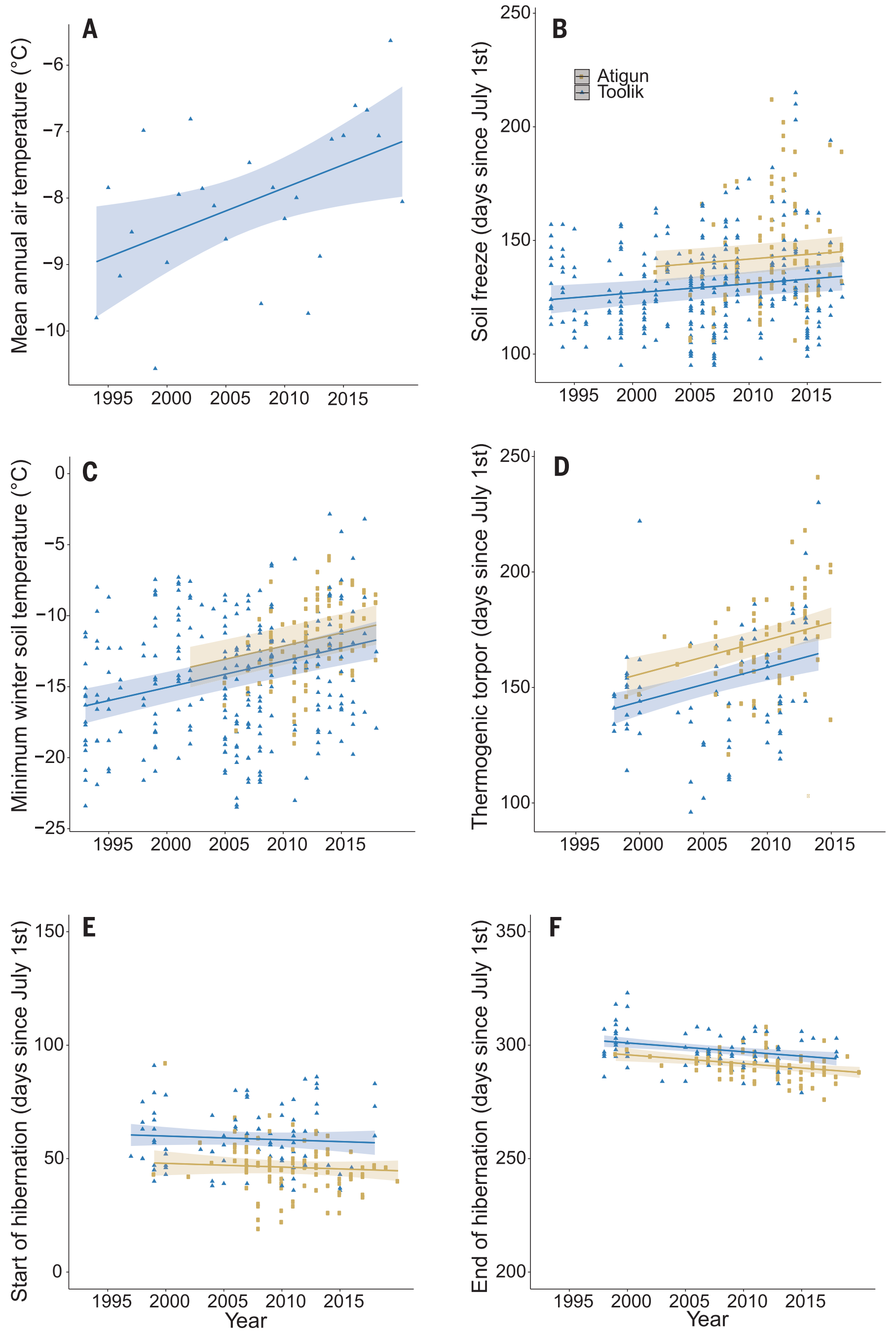


Fig. 2. Long-term changes in temperature and hibernation physiology. Over the past 25 years at Toolik Field Station in the Alaskan Arctic, mean annual air temperatures increased (**A**), soil freeze date occurred later (sampling at 1-m depth, adjacent to hibernacula) (**B**), and minimum overwinter soil temperatures increased (**C**). Concurrently, the onset of thermogenic torpor occurred later in the fall [(**D**); shown for females only]. Although the date at which female ground squirrels initiated hibernation in the fall did not change over time (**E**), the date that they ended hibernation in the spring became earlier (**F**). Lines represent mean results of linear mixed effects models, and shaded regions represent 95% confidence intervals of the mean. Points represent raw data, and colors represent study sites (Atigun or Toolik). For all animal data, data points are plotted against the year in which an individual initiated hibernation.

($n = 262$, $\beta = -0.36$, 95% credible interval, −0.57 to −0.15). In combination, this resulted in an ~10-day reduction in the annual duration that soil was frozen at 1-m depth over the study period.

In arctic ground squirrels, the date on which core body temperatures first decreased below 0°C each hibernation season, i.e., the date when squirrels first became thermogenic during torpor (*9*), was delayed by ~15 days per decade over the 25-year course of our study ($n = 256$, $\beta = 1.484$, $t = 4.800$, df = 91, $P < 0.001$; Fig. 2D), with no difference seen between sexes ($\beta = 3.039$, $t = 1.041$, df = 161, $P = 0.299$). However, other phenological events did not show the same pattern. Neither adult males ($n = 122$, $\beta = -0.354$, $t = -1.454$, df = 32, $P = 0.156$) nor adult females ($n = 182$, $\beta = -0.163$, $t = -0.819$, df = 71, $P = 0.416$; Fig. 2E) changed the date at which hibernation began in fall. Adult females, however, ended hibernation earlier in spring ($n = 166$, $\beta = -0.389$, $t = -3.592$, df = 60, $P < 0.001$; Fig. 2F) at a rate of ~4 days/decade. As females ended hibernation earlier over the 25-year period, animal mass measured during the first 2 weeks after emergence increased ($n = 91$, $\beta = 3.472$, 95% credible interval 0.196 to 6.781; see the supplementary materials and fig. S3). Males did not shift spring phenology ($n = 120$, $\beta = -0.192$, $t = -1.103$, df = 30, $P = 0.279$). No directional changes were observed in parturition timing ($n = 91$, $\beta = -0.003$, $t = -1.343$, df = 28, $P = 0.190$).

Discussion

Winter plays a fundamental role in determining species' range limits and local population dynamics (*6*, *16*). We show that warming in the AlaskanArctic has altered seasonal freeze-thaw dynamics of the soil active layer and thus the thermoregulatory patterns of arctic ground squirrels during hibernation. Delayed freeze of the soil active layer near hibernacula is delaying the onset of thermogenic torpor and presumably reducing energy expenditure and overwinter weight loss. Further, we found sex differences in phenological responses to climate change, with females advancing their spring active season by 10 days over 25 years and males showing no change.

Warmer winter conditions and the shortening of the hibernation season in females have the potential to affect the survival probability of free-living arctic ground squirrels. There are several general mechanisms through which this could occur, including changes in energetics and predation exposure. First, delaying thermogenic torpor will decrease the total amount of time that arctic ground squirrels spend defending body temperature during hibernation, which will reduce overwinter energy expenditures (*9*, *17*) and increase spring mass in females, as we have demonstrated in this study. Additionally, changes in winter temperatures will alter the energetic costs of torpor bouts and episodic interbout arousals (*8*) that characterize hibernation in small mammals. Arctic ground squirrels spend most of the hibernation season generating heat to avoid freezing, unlike most hibernators in more moderate climates, which can safely thermoconform because the surrounding soil temperatures remain above freezing during torpor (*18*). Reduced intensity and/or duration of thermogenesis caused by warmer conditions would

allow arctic ground squirrels to conserve energy and potentially increase winter survival, particularly among vulnerable, energy-limited age classes such as juveniles (*19*, *20*). Alternatively, warmer conditions associated with climate change could shorten the hibernation season and increase the number of days that arctic ground squirrels are active above ground, increasing mortality rates because of increased exposure to predators (*21*). Thus, the consequences of climate change for hibernators will likely be heterogeneous (*22*). Although our study focused on a system in which warmer air temperatures lead to warmer hibernacula temperatures and likely reduced energy expenditure, in other systems, climate change may reduce snow depth, which can decrease burrow temperatures and decrease energetic savings through hibernation (*20*).

We found sex differences in phenological flexibility, with female arctic ground squirrels, but not males, terminating hibernation earlier. Other studies suggest that female ground squirrels are less sensitive to the direct effects of temperature (*23*, *24*) and instead are responsive to temperature-driven changes in spring snow cover conditions (*25*). Thus, female flexibility appears to allow them to match energetic demands with the environment, and we expect that earlier snowmelt will also correspond with earlier vegetation green-up (*26*). As winters continue to warm, sex differences in phenological shifts may lead to disrupted intersexual interactions. For example, during one extremely warm spring in eastern Canada, female Richardson's ground squirrels (*Urocitellus richardsonii*) became sexually receptive before most males were physiologically prepared to mate, resulting in a phenological mismatch between the sexes (*27*). In the short term, mismatches such as these could affect population reproductive rates, and over longer time scales, continued warming in the Arctic may be a strong selective force resulting in evolutionary changes in male phenology.

The consequences of climate change may be direct, such as the changes in hibernation physiology and phenology that we report here, as well as indirect. If the population size and temporal availability of arctic ground squirrels above ground are altered by climate change, then this could have cascading indirect effects on diverse tundra predators. Understanding the impact of climate change on species such as the arctic ground squirrel will aid assessments of how arctic food webs will function in a rapidly warming world, and research linking physiology and phenology to demographic responses is an important component of understanding community responses to climate change.

REFERENCES AND NOTES

1. F. Pithan, T. Mauritsen, *Nat. Geosci.* **7**, 181–184 (2014).
2. J. E. Box *et al.*, *Environ. Res. Lett.* **14**, 045010 (2019).
3. J. E. Hobbie *et al.*, *Ambio* **46** (Suppl. 1), 160–173 (2017).
4. E. Post, M. C. Forchhammer, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 2369–2375 (2008).
5. G. Gauthier *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120482 (2013).
6. C. M. Williams, H. A. L. Henry, B. J. Sinclair, *Biol. Rev. Camb. Philos. Soc.* **90**, 214–235 (2015).
7. K. Wilsterman, M. A. Ballinger, C. M. Williams, *Funct. Ecol.* **35**, 11–31 (2021).
8. M. M. Richter *et al.*, *Physiol. Biochem. Zool.* **88**, 81–89 (2015).
9. B. M. Barnes, *Science* **244**, 1593–1595 (1989).
10. S. P. Caro, S. V. Schaper, R. A. Hut, G. F. Ball, M. E. Visser, *PLOS Biol.* **11**, e1001517 (2013).
11. H. E. Chmura, C. T. Williams, *Horm. Behav.* **144**, 105215 (2022).
12. S. J. Thackeray *et al.*, *Glob. Chang. Biol.* **19**, 3568–3580 (2013).
13. M. D. Burgess *et al.*, *Nat. Ecol. Evol.* **2**, 970–975 (2018).
14. C. L. Buck, B. M. Barnes, *J. Mammal.* **80**, 1264–1276 (1999).
15. C. T. Williams, B. M. Barnes, C. L. Buck, *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **202**, 53–62 (2016).
16. D. Gudex-Cross *et al.*, *Glob. Ecol. Biogeogr.* **31**, 1366–1380 (2022).
17. S. A. Karpovich, Ø. Tøien, C. L. Buck, B. M. Barnes, *J. Comp. Physiol. B* **179**, 691–700 (2009).
18. T. Ruf, F. Geiser, *Biol. Rev. Camb. Philos. Soc.* **90**, 891–926 (2015).
19. C. Lenihan, D. Van Vuren, *Can. J. Zool.* **74**, 297–302 (1996).
20. C. Rézouki *et al.*, *J. Anim. Ecol.* **85**, 761–773 (2016).
21. C. Turbill, S. Prior, *Funct. Ecol.* **30**, 1366–1372 (2016).
22. C. P. Wells, R. Barbier, S. Nelson, R. Kanaziz, L. M. Aubry, *Ecography* **2022**, e06056 (2022).
23. B. M. Barnes, D. Ritter, in *Life in the Cold: Ecological, Physiological, and Molecular Mechanisms* (CRC Press, 1993), pp. 119–130.
24. H. E. Chmura *et al.*, *Integr. Comp. Biol.* **62**, 1012–1021 (2022).
25. C. T. Williams *et al.*, *Am. Nat.* **190**, 854–859 (2017).
26. I. P. La Puma, T. E. Philippi, S. F. Oberbauer, *Remote Sens. Environ.* **109**, 225–236 (2007).
27. C. E. Kucheravy *et al.*, *Sci. Rep.* **11**, 21684 (2021).
28. Data for: H. E. Chmura, C. Duncan, G. Burrell, B. M. Barnes, C. L. Buck, C. T. Williams, Climate change is altering the physiology and phenology of an arctic hibernator, Dryad (2023); https://doi.org/10.5061/dryad.pzgmsbcqq.

ACKNOWLEDGMENTS

We thank J. T. Moore, F. Kohl, M. Richter, and many additional students and technicians who worked countless hours in the field to support this work. We are also grateful for 30 years of outstanding support from staff and facilities at the Toolik Field Station, which made this research possible. **Funding:** This work was supported by the National Science Foundation (grant IOS-1558056 to C.T.W. and C.L.B.); the National Science Foundation (grant IOS-1558160 to B.M.B.); a University of Alaska Centennial Postdoctoral Fellowship (H.E.C.); the National Institute of General Medical Sciences of the National Institutes of Health [Institutional Development Award (IDeA) grant P20GM103395 to B.M.B. and C.D.]; and a University of Alaska Fairbanks URSA undergraduate research award (G.B.). **Author contributions:** Conceptualization: H.E.C., C.T.W.; Data curation: H.E.C., C.D., G.B.; Formal analysis: H.E.C., C.D., G.B.; Funding acquisition: B.M.B., C.L.B., C.T.W.; Investigation: H.E.C., C.D., G.B., C.L.B., C.T.W.; Methodology: H.E.C., B.M.B., C.L.B., C.T.W.; Project administration: C.T.W.; Resources: B.M.B., C.L.B., C.T.W.; Supervision: C.T.W.; Visualization: H.E.C.; Writing – original draft: H.E.C., C.T.W.; Writing – review and editing: H.E.C., B.M.B., C.L.B., C.T.W. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The datasets and code generated during the study are available in the Dryad repository (*28*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf5341
Materials and Methods
Supplementary Text
Figs. S1 to S3
Table S1
References (*29*–*39*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 31 October 2022; accepted 25 April 2023
10.1126/science.adf5341

HUMAN GENETICS

On the genes, genealogies, and geographies of Quebec

Luke Anderson-Trocmé[1,2], Dominic Nelson[1,2], Shadi Zabad[3], Alex Diaz-Papkovich[1,4], Ivan Kryukov[1,2], Nikolas Baya[5], Mathilde Touvier[6], Ben Jeffery[5], Christian Dina[7], Hélène Vézina[8], Jerome Kelleher[5], Simon Gravel[1,2]*

Population genetic models only provide coarse representations of real-world ancestry. We used a pedigree compiled from 4 million parish records and genotype data from 2276 French and 20,451 French Canadian individuals to finely model and trace French Canadian ancestry through space and time. The loss of ancestral French population structure and the appearance of spatial and regional structure highlights a wide range of population expansion models. Geographic features shaped migrations, and we find enrichments for migration, genetic, and genealogical relatedness patterns within river networks across regions of Quebec. Finally, we provide a freely accessible simulated whole-genome sequence dataset with spatiotemporal metadata for 1,426,749 individuals reflecting intricate French Canadian population structure. Such realistic population-scale simulations provide opportunities to investigate population genetics at an unprecedented resolution.

The tapestry of human genetic history is formed of ancestral lineages interwoven by generations of coalescence and recombination events (*1*). It was woven across geographic landscapes (*2*) by individual dispersal and historical waves of migrations that can sometimes be reconstructed by genomic analyses (*3*, *4*). Yet the complex relationship between spatial migrations and genetic variation still poses formidable challenges (*5*, *6*).

As a general trend, the limits of dispersal lead to continuous isolation by distance and a correlation between genetic and geographic distances (*7*–*9*). However, in any region, specific historical events or geographic barriers are often used to explain discrete patterns of population variation (*5*, *10*, *11*). Reconciling continuous variation into discrete "evolutionarily significant units" has proven to be difficult and sometimes misleading (*12*–*14*). While many studies have considered anisotropic migration models (*15*), and even detailed models of geographic constraint or "resistance" (*16*, *17*), comparing these models to genetic data is challenging.

This study takes advantage of a population-scale spatially labeled pedigree (or spatial pedigree) compiled from more than 4 million Catholic parish records in the province of Quebec. Using genotype data for 20,451 individuals and newly created pedigree-aware simulation tools, we provide a detailed spatiotemporal model of genetic variation at scales ranging from tens to thousands of kilometers. By including French and British individuals in our analyses, we assess how much ancestral population structure has been preserved from these two populations. We highlight the relationship between river networks and genetic similarity as the past four centuries of European colonial history has been marked by rapid frontier expansion, beginning along the shores of the St. Lawrence River and eventually expanding up its tributaries. By tracing the genealogical ancestry of millions of individuals across space and time, we describe a constellation of distinct founder events arranged along geographic features that defined transportation and economic activity. In doing so, we further bridge the gaps between family pedigrees and continental population structure as well as gaps between theoretical models and empirical demographic histories.

Results

Regional distribution of genetic variation

Quebec, a province in Canada, has a population of 8.8 million individuals, of which ~6.5 million speak French as a primary language. Most individuals in Quebec derive ancestry from ~8500 settlers who migrated from France in the 17th and 18th centuries—we refer to these individuals as French Canadians. The 2600 French settlers contributed two-thirds of the French Canadian gene pool (*18*). French settlers occupied territory inhabited and used by First Nations for thousands of years (*19*). Despite folk histories implying large amounts of Indigenous ancestry among French Canadians (*20*), genetic and genealogical studies show that French Canadians born in Quebec carry on average <1% of ancestry tracing back to Indigenous populations, with the rest being mostly attributed to French ancestry (*21*).

Catholic marriage records in Quebec have been digitized and combined with civil registry data by the BALSAC Project (*22*). Because of a strong historical correlation between French ancestry and Catholic religion, the pedigree reconstructed from these records is particularly complete among French Canadians (*22*). We therefore focused our genetic and genealogical analyses on 20,451 individuals inferred to be French Canadian from 21,068 individuals among the CARTaGENE [12,064 (*23*)] and Genizon cohorts [9004, first reported here; see (*24*) for details on cohorts and ancestry inference]. We used principal components analysis (PCA, fig. S1) and uniform manifold approximation and projection (UMAP, Fig. 1A) (*25*, *26*) to visualize genetic variation in the population. A schematic summary of the entire French Canadian spatial pedigree and the geographic location of 4882 individuals linked to the spatial pedigree are shown in Fig. 1, B and C. Visual inspection shows strong correlation of genetic and spatial proximity and suggests that gradients in genetic variation coincide with geographical barriers and conduits such as the St. Lawrence, Saguenay, and Chaudière rivers or the Laurentian and Appalachian mountains (see tables S1 and S2 for glossaries of placenames in Quebec and France).

French ancestry uprooted

Analyses of genealogical records show that most French settlers arrived in Quebec City and came from regions in western (Aunis, Poitou) and northwestern (Normandy, Perche) France, as well as densely populated Ile-de-France [see (*27*, *28*) and table S2]. However, successive waves of migration related to military and

Table 1. Proportion of founder events shared between regions, as measured by the ratio of the cross-coalescence rate divided by the within-region coalescence rate.

	Proportion of founder events shared with:		
Region	SLSJ	Bas-Saint-Laurent	Beauce
Saguenay–Lac-Saint-Jean (SLSJ)	–	0.055	0.035
Bas-Saint-Laurent	0.456	–	0.358
Beauce	0.412	0.516	–

[1]Department of Human Genetics, McGill University, Montreal, QC, Canada. [2]McGill University Genome Centre, Montreal, QC, Canada. [3]School of Computer Science, McGill University, Montreal, QC, Canada. [4]Quantitative Life Sciences, McGill University, Montreal, QC, Canada. [5]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [6]Sorbonne Paris Nord University, INSERM U1153, INRAE U1125, CNAM, Nutritional Epidemiology Research Team (EREN), Epidemiology and Statistics Research Center, University Paris Cité (CRESS), Bobigny, France.
[7]Nantes Université, CNRS, INSERM, l'institut du thorax, Nantes, France. [8]BALSAC Project, Université du Québec á Chicoutimi, Chicoutimi, QC, Canada.
*Corresponding author. Email: simon.gravel@mcgill.ca

colonization objectives had different origins and demographics (*28*). To assess whether this history is reflected in the Quebec population structure, we compared the genomes of individuals living in different regions of Quebec and France. In agreement with historical records, French Canadians share more-recent ancestry with individuals from western France as measured by DNA that is identical by descent (IBD) (fig. S2A). We also observe more-modest variation across Quebec regions in their IBD with France, with regions of Saguenay–Lac-Saint-Jean (SLSJ), Beauce, and Bas-Saint-Laurent having slightly elevated values, likely because of founder events (see fig. S2B and discussion in the supplementary materials). We used F_4-statistics to assess whether genetic differentiation between any pair of Quebec regions correlates with genetic differentiation between any pair of European regions (seven French regions and Britain) and found no such correlation (fig. S3). This suggests that most present-day structure among French Canadians is independent of ancestral structure or differential contributions by French and British founders at this broad scale.

Simulated genomes with known transmission histories

To assess how much of the French Canadian population structure can be accounted for by events after the arrival of French settlers, we generalized the msprime software (*29*–*31*) to perform genome-wide coalescent simulations conditioning on the known pedigree of the French Canadian population. We defined the FixedPedigree ancestry simulation model in msprime version 1.2 to trace the ancestry of samples back through the genealogy accounting for coalescence and recombination (fig. S4). To account for relatedness beyond the founders of the known pedigree, we modeled coalescence under a demographic model for European ancestry (*32*) [see fig. S5 and (*24*) for details].

We compared the simulations to ascertained data with the subset of 4882 individuals who were both genotyped and linked to the pedigree. While previous work had shown qualitative concordance between genealogical and genetic distances (*33*), here the PCA and UMAP of the simulations show clear quantitative agreement with the fine-scale structure of ascertained genotype data, as is evident from the first six PCs showing strong correlation (Fig. 2 and fig. S6). We performed 10 replicate simulations and find little variance in their concordance to real data (fig. S7A). When comparing

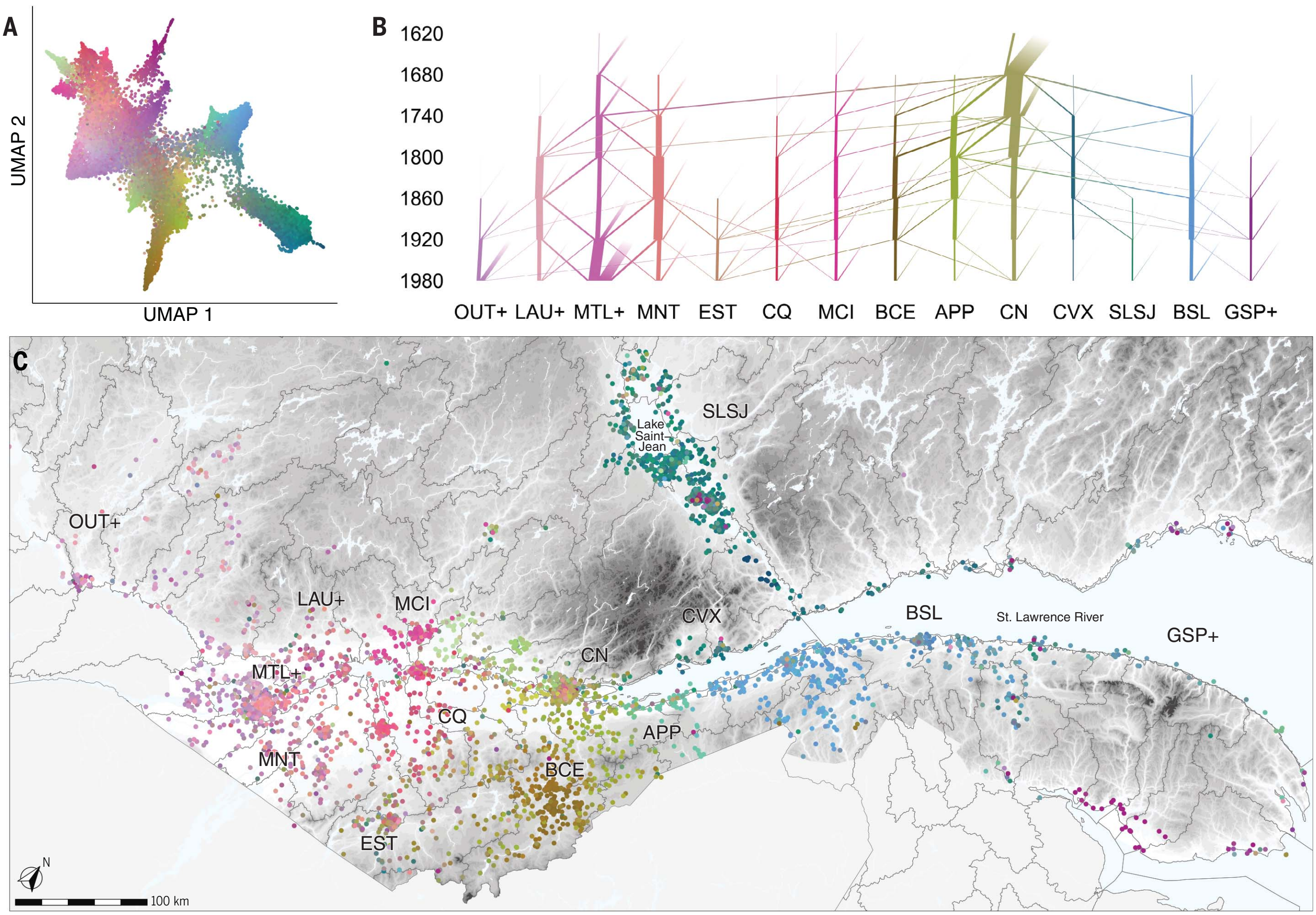


Fig. 1. French Canadian genes and genealogies mirror Quebec's geographies. (**A**) UMAP for 20,451 individuals with inferred French Canadian ancestry. Each individual is assigned a color on the basis of their location in a dimensionally reduced genetic projection [see (*24*)]. (**B**) Visualizing the ancestry of French Canadians across regions (*x* axis) over time of marriage (*y* axis). The thickness of the line at time *t* from location *A* to location *B* represents the expected amount of genetic material ancestral to the present-day population, estimated from the genealogical data. Fading lines indicate genealogical founders (i.e., individuals without documented parents). The thick fading line at the top indicates the arrival of French settlers to Quebec City in the Capitale-Nationale (CN) region. (**C**) Clusters in genetic space coincide with distinct geographic regions of Quebec defined by geographic features such as rivers and mountains. Watershed boundaries are shown in black. OUT+, Outaouais+; LAU+, Laurentides+; MTL+, Montréal+; MNT, Montérégie; EST, Estrie; CQ, Centre-du-Québec; MCI, Mauricie; BCE, Beauce; APP, Appalaches; CVX, Charlevoix; SLSJ, Saguenay–Lac-Saint-Jean; BSL, Bas-Saint-Laurent; GSP+, Gaspésie+. See table S1 and fig. S21 for place-name definitions.

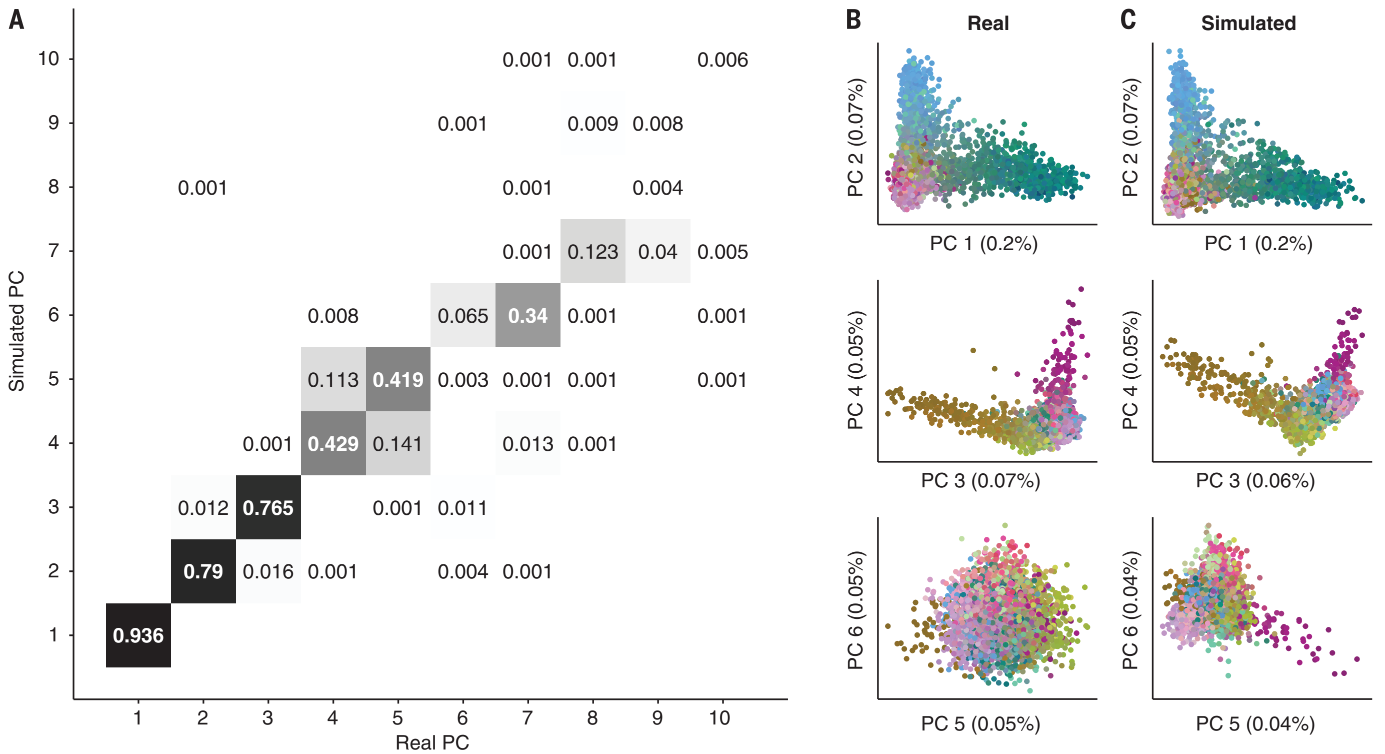


Fig. 2. Simulated genomes capture observed population structure. Comparison of the same 4882 individuals using observed and simulated genomes (colored as in Fig. 1; see supplementary materials for details). (**A**) The correlation between observed and simulated principal components. Values different from 0.000 are shown. (**B**) PCA projections of observed genomes. (**C**) PCA projections of simulated genomes.

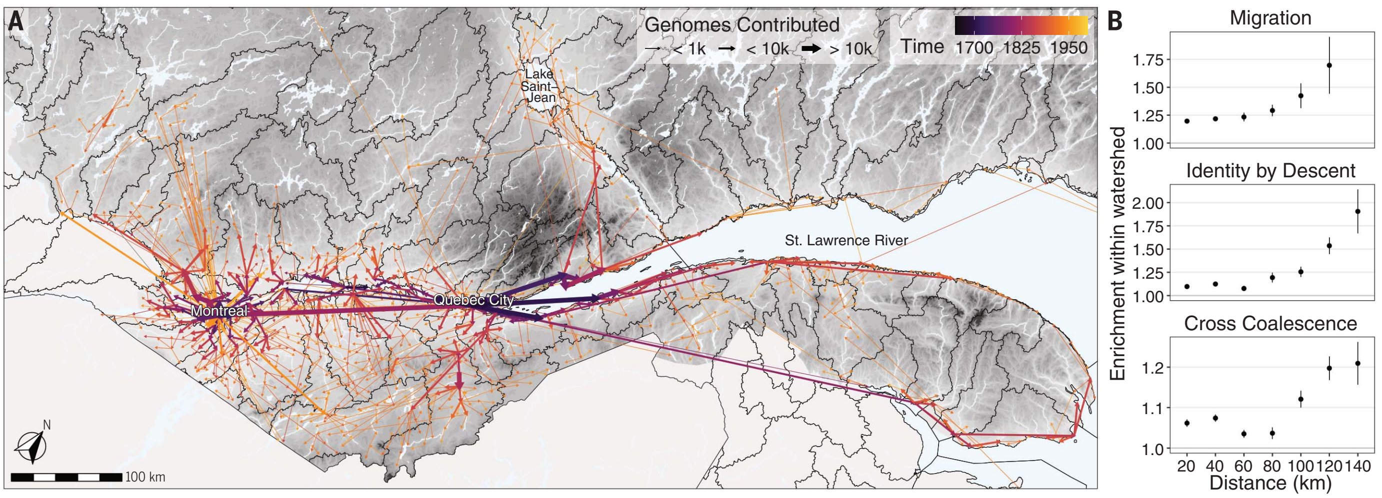


Fig. 3. Watersheds influence French Canadian migrations and relatedness. (**A**) Primary routes of estimated genetic ancestry. Segments link each town to the town from which migrants are estimated, on the basis of the genealogy, to have contributed the most genetic material. The width of segments indicates the total expected genetic contribution to present-day individuals, and the color indicates the mean date of when these expected contributions occurred historically. To avoid overplotting, we excluded the region of Abitibi-Témiscamingue and migrations of <10 km and migrations contributing <10 genomes. (**B**) Relatedness is enriched within watersheds. Black dots indicate the excess migrations and relatedness for towns within watersheds relative to towns in different watersheds at a fixed distance (*24*).

replicate simulations to each other, we find high correlations for the leading eight PCs and a drop-off of correlations for higher-order PCs: the randomness of mendelian inheritance in this fixed pedigree has a small effect on leading PCs but perturbs or reorders higher-order principal components (fig. S7B). Ultimately, leading axes of genetic variation among French Canadians reflect genetic drift that followed French settlement and is encoded in the spatial pedigree.

Given the accuracy of the simulated data, we simulated whole genomes of 1.4 million present-day individuals with four grandparents linked to the pedigree to use as a resource in genetic modeling. Tree sequences of these simulations,

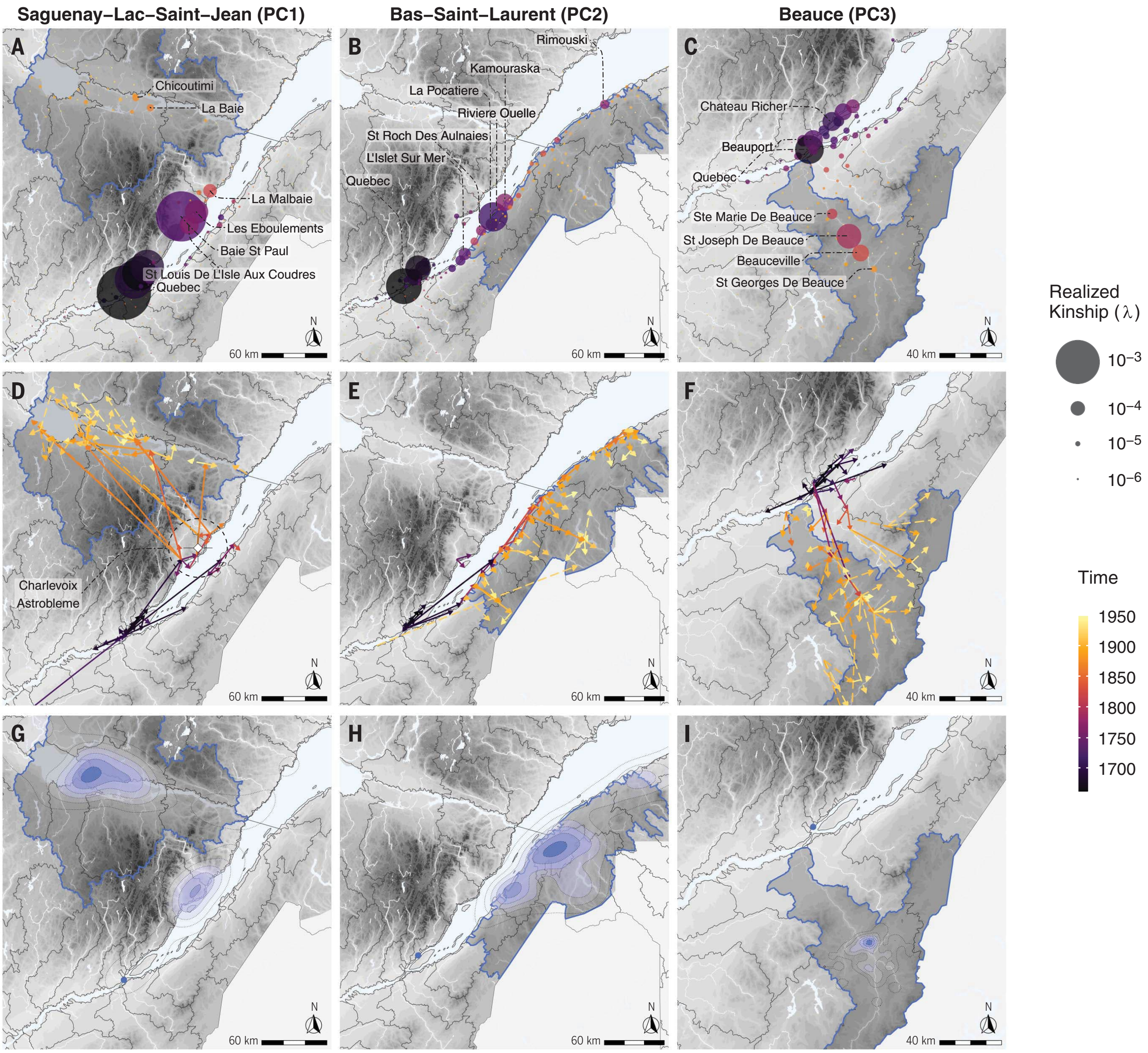


Fig. 4. Historical migrations and founder events define population structure. (**A** to **C**) Time, location, and stringency of founder events in three regions of Quebec defined by the top three principal components (fig. S13). Size of points is proportional to realized kinship for the present-day French Canadian population, estimated from the genealogy. Point color represents mean date of realized kinship based on marriage dates of ancestors. (**D** to **F**) Major axes of migration, as measured by the estimated genetic contributions to present-day individuals living within the highlighted regions. Dotted arrows indicate towns having the largest genetic contribution to towns within the highlighted regions, and solid arrows highlight the fifth percentile of migratory routes ranked by estimated genetic contribution. Arrow colors represent the mean time of the genetic contribution based on the marriage dates of ancestors. (**G** to **I**) The dispersal range for the single ancestor with the highest contribution to each of the three regions. For a given ancestor, we generated a heatmap using the location of towns weighed by the average contribution of that ancestor to the probands of each town. The marriage location of each of the major contributors is indicated by a blue dot.

including decade and spatial information on all recorded ancestors, are freely available on Zenodo (*34*). UMAP and PCA analyses for the 1.4 million individuals are presented in fig. S8, showing a continuous triangular distribution of points along the first two principal components (PCs), as in Fig. 2C, but a star-like distribution for the next 16 PCs, with opposite star branches corresponding to fine-scale differentiation between neighboring regions sometimes 20 km apart (fig. S9).

Gene flow within watershed boundaries

Figure 1 and fig. S1 suggest that genetic similarity varies continuously over space, in a manner reminiscent of population structure across Europe (*9*). However, each PC tends to differentiate a specific region rather than a broad geographic cline, in a manner more reminiscent of finer-scale differentiation across alpine valleys (*5*). To assess how migration rates are affected by topographical features such as rivers and mountains, we visualized the major migration events shaping the French Canadian population through space and time. Figure 3A exhibits waves of frontier expansion consisting of a series of sequential migrations up the

tributaries of the St. Lawrence River, ranging from tens to hundreds of kilometers in scale.

The first permanent French settlement took its name from the Algonquin word *kebec*, referring to the region where the St. Lawrence River becomes narrow (*35*). Quebec City was founded at a strategic bridgehead location for the French as they sought to gain control of the main entrance to the Great Lakes in the 17th century [(*36*), pp. 49–52]. Facing vast forested territory used and occupied by Iroquoian and Woodland First Nations (*19*), the French formed a fluvial colony with thin ribbons of settlements along shorelines using a riverfront land division strategy [(*36*), pp. 56–57].

To quantify patterns of relatedness and dispersal among individuals distributed in 1698 parishes across the landscape, we considered three measures of relatedness and migration propensity between distinct geographic locations: migration rates, identity by descent, and genealogical estimates of cross-coalescence rates. All three show clear patterns of isolation by distance in all regions (figs. S10 and S11). Given the importance of rivers in early settlement strategies, economic activity, and transportation (*37*), we hypothesized that relatedness and migration patterns broadly followed directions defined by local rivers and are therefore enriched for towns within watersheds. For each of the three metrics, we computed this enrichment for a tessellating set of 80 watersheds as a function of distance (Fig. 3B) (*24*). We find an enrichment for the three metrics for all distances. At 120 km, where the enrichment is strongest, we find a 40% enrichment in migration rates, a 75% enrichment in genetic relatedness rates, and a 20% enrichment in cross-coalescence rates. However, each region separately has a distinct migration pattern with varying degrees of topographical influence (Fig. 1C and fig. S12).

Historical migrations in space and time

To follow the formation of regional substructure in the French Canadian population, we defined three regions on the basis of watersheds with large proportions of individuals driving the top principal components (fig. S13). These empirical regions closely align with the SLSJ, Bas-Saint-Laurent, and Beauce regions defined above, so we reuse these labels to refer to the watershed-based empirical regions. We ascended the spatial pedigree for pairs of individuals living in each region, computed the estimated realized kinship (equivalently, coalescence) for each common ancestor (*24*), and determined the location, timing, and stringency of founder events (Fig. 4, A to C). We use the term “founder event” to refer to place-times where many coalescences are expected to have occurred (i.e., many pairs of individuals found a common ancestor), corresponding to a small effective population size but not necessarily a small census size.

All regions exhibit their strongest founder event in Quebec City but have distinct patterns of subsequent differentiation. SLSJ [realized kinship (λ) = 0.005; Fig. 4, A, D, and G] has a dominant early founder event in Baie-Saint-Paul (λ = 0.001) and neighboring towns in Charlevoix, where an astrobleme—a geological formation resulting from the erosion of an impact crater—created a small pocket of fertile land within otherwise mountainous terrain (fig. S14) (*38*). Limited carrying capacity within the astrobleme led to demographic pressure and subsequent rapid expansion up the Saguenay River [(*36*), p. 91], resulting in a vast majority of kinship predating the colonization of SLSJ. The Beauce region (λ = 0.002; Fig. 4, C, F, and I) has a handful of founder events in Saint-Joseph-De-Beauce (λ = 0.0003) and along the Chaudière River, with migrations reminiscent of a hub-and-spoke model. Finally, the Bas-Saint-Laurent (λ =0.002; Fig. 4, B, E, and H) has an assortment of founder events including a dominant founder event in Rivière-Ouelle (λ = 0.0004) but also more minor founder events scattered across hundreds of kilometers of shoreline, acting as a one-dimensional regional hub for subsequent inland migrations (Fig. 4E).

As expected in an expanding population, some early settlers (those identified as superfounders) made a large contribution to the present-day population (*39*, *40*). The top 10 superfounders in each region contributed 37, 12, and 14% of the realized kinship in SLSJ, Bas-Saint-Laurent, and Beauce, respectively (figs. S15 to S17). Surprisingly, while each region has its oldest and strongest founder event near Quebec City, no two regions share the same superfounders (fig. S18). To assess the overlap of each regional founder event, we computed cross-coalescence rates (*41*) [see (*24*) and fig. S19] and find that 35 to 50% of kinship in Bas-Saint-Laurent and Beauce can be attributed to a shared founder event (Table 1), but the founder event in SLSJ is only 5% shared, reflecting more-intense founder events in Charlevoix and the region around Quebec City that are specific to SLSJ.

Not all regions of Quebec exhibit such spatially defined founder events. Even though Abitibi-Témiscamingue was settled by a process of rapid frontier expansion similar to that seen in SLSJ, it did not lead to founder events reflected in leading axes of genetic differentiation. In contrast to the events in SLSJ, French Canadian settlers to Abitibi-Témiscamingue came from numerous villages scattered throughout the province (fig. S20). While many of the villages in Abitibi-Témiscamingue have measurable founder events, these seldom overlap at a regional level. To illustrate this, we consider the villages of Rémigny and Rollet separated by 20 km along the Ottawa River (fig. S20). Cross-coalescence rates of these villages have 11% overlap. For comparison, La Baie and Roberval in SLSJ have 70% overlap despite being >100 km apart. Thus, Abitibi-Témiscamingue shows limited evidence of a shared founder event and almost no isolation by distance (fig. S10), yet the ancestral structure, together with parallel founding events, creates important substructure (fig. S20). Such cryptic substructure may be a common feature of human evolution (*42*).

Discussion

Classical population genetic models often approximate reproduction and mate selection as a uniform random process. By approximating the effects of a myriad of individual motivations and choices that are unavailable to scientists, classical models provide an explanation for trends such as drift and selection. However, recent work has highlighted the benefits of a more-detailed genealogical model (*4*, *43*). As we give up the simplified assumption of uniform random mating, the number of demographic parameters relevant to evolution grows rapidly.

The BALSAC pedigree—a particularly complete population-scale spatial pedigree—has been instrumental in identifying multigenerational demographic effects such as the reproductive advantages of being on a wave-front expansion (*39*) or the transmissibility of family size (*44*) and migration propensity (*45*). Others have used it to study broad-scale variation in genetic and genealogical relatedness (*33*, *46*) and its historical determinants, such as the kinship of the first settlers of Charlevoix (*37*) or the delay in the settlement of the Saguenay owing to restrictions related to the fur trade [(*36*), p. 91].

In this study, we sought to develop a comprehensive genetic model that captured all these effects and more. We used the concept of realized kinship (or coalescence) to break down founder events as a sum of expected contributions from historical individuals and towns. We used this model to highlight how one of the best-studied human founder populations in SLSJ was influenced by the distinctive geography of the region that was shaped by a cosmic event occurring 400 million years ago in Charlevoix. By providing a first detailed description of other founder events in Quebec, we showed a wide diversity of founder dynamics. We also found concordant genetic and genealogical support for the idea that geographic features such as rivers and mountains played a systematic role in defining major axes of migration and genetic variation. Finally, the strong correlation between empirical and simulated genetic data provides evidence that the structure within the French Canadian population can largely be attributed to events in North America, while the population kept a genetic signature of the regions in France that contributed more early French settlers.

We described how idiosyncratic events as well as geological, social, and historical effects translated to genetic variation patterns at various

geographic scales and over centuries. Although our simulations are based on a real pedigree, they do not contain identifying information, and we can freely share this genome-wide dataset along with spatiotemporal metadata for >1.4 million individuals. Geneticists often rely on simulations to assess accuracy and robustness of methods for genetic risk prediction, historical inference, and genealogical inference. For this reason, much work has been done building simulation models that more accurately capture genealogical structure (*30*, *47*, *48*). However, these rely on strong demographic assumptions such as random mating. We believe that a large and freely available simulated dataset with realistic population structure and mating patterns will help the design of more-robust inference methods.

Of course, these simulations are far from perfect. They do not account for natural selection beyond what is captured in the pedigree (*39*) or for any structure among the founders. The pedigree itself contains some amount (likely <1%) of recording errors (*49*). Such errors seem to have a limited effect on broad-scale population structure, but they will continue to pose challenges for finer-scale genealogical analyses (*43*, *49*). Despite these caveats, we believe that the genetic model, the simulation tools, and the publicly available simulated data we describe here provide a lens to investigate population genetics at an unprecedented resolution.

REFERENCES AND NOTES

1. A. W. Wohns *et al*., *Science* **375**, eabi8264 (2022).
2. L. L. Cavalli-Sforza, A. Piazza, P. Menozzi, J. Mountain, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6002–6006 (1988).
3. B. M. Henn, L. L. Cavalli-Sforza, M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17758–17764 (2012).
4. G. S. Bradburd, P. L. Ralph, *Annu. Rev. Ecol. Evol. Syst.* **50**, 427–449 (2019).
5. L. L. Cavalli-Sforza, *Sci. Am.* **221**, 30–37 (1969).
6. C. J. Battey, P. L. Ralph, A. D. Kern, *Genetics* **215**, 193–214 (2020).
7. S. Wright, *Genetics* **28**, 114–138 (1943).
8. S. Ramachandran *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15942–15947 (2005).
9. J. Novembre *et al*., *Nature* **456**, 98–101 (2008).
10. C. Bycroft *et al*., *Nat. Commun.* **10**, 551 (2019).
11. A. Saint Pierre *et al*., *Eur. J. Hum. Genet.* **28**, 853–865 (2020).
12. N. A. Rosenberg *et al*., *PLOS Genet.* **1**, e70 (2005).
13. G. S. Bradburd, G. M. Coop, P. L. Ralph, *Genetics* **210**, 33–52 (2018).
14. C. Moritz, *Trends Ecol. Evol.* **9**, 373–375 (1994).
15. F. Jay, P. Sjödin, M. Jakobsson, M. G. Blum, *Mol. Biol. Evol.* **30**, 513–525 (2013).
16. D. Petkova, J. Novembre, M. Stephens, *Nat. Genet.* **48**, 94–100 (2016).
17. B. H. McRae, *Evolution* **60**, 1551–1561 (2006).
18. H. Charbonneau *et al*., *The First French Canadians: Pioneers in the St. Lawrence Valley* (Univ. of Delaware Press, 1993).
19. Aboriginal and Northern Affairs Canada, *First Nations in Canada* (2013); https://www.rcaanc-cirnac.gc.ca/eng/1307460755710/1536862806124#chp0.
20. Y. Beauregard, *Cap-aux-Diamants* **34**, 38–42 (1993).
21. C. Moreau *et al*., *PLOS ONE* **8**, e65507 (2013).
22. H. Vézina, J.-S. Bournival, *Hist. Life Course Stud.* **9**, 114–129 (2020).
23. P. Awadalla *et al*., *Int. J. Epidemiol.* **42**, 1285–1299 (2013).
24. Materials and methods are available as supplementary materials.
25. L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction arXiv:1802.03426 [stat.ML] (2018).
26. A. Diaz-Papkovich, L. Anderson-Trocmé, C. Ben-Eghan, S. Gravel, *PLOS Genet.* **15**, e1008432 (2019).
27. S.-A. Lortie, *Bulletin du parler Français au Canada* **2**, 17–18 (1903).
28. H. Vézina, M. Tremblay, B. Desjardins, L. Houde, *Cah. Que. Demogr.* **34**, 235–258 (2005).
29. J. Kelleher, A. M. Etheridge, G. McVean, *PLOS Comput. Biol.* **12**, e1004842 (2016).
30. D. Nelson *et al*., *PLOS Genet.* **16**, e1008619 (2020).
31. F. Baumdicker *et al*., *Genetics* **220**, iyab229 (2022).
32. J. A. Tennessen *et al*., *Science* **337**, 64–69 (2012).
33. M.-H. Roy-Gagnon *et al*., *Hum. Genet.* **129**, 521–531 (2011).
34. L. Anderson-Trocmé, Simulated genomes from manuscript “On the Genes, Genealogies and Geographies of Quebec”, version 1.0, Zenodo (2023); https://doi.org/10.5281/zenodo.7702392.
35. Y. Hébert, *Les Ponts de glace sur le Saint-Laurent* (GID, 2012).
36. S. Courville, *Quebec: A Historical Geography* (UBC Press, 2009).
37. G. Bouchard, M. De Braekeleer, *Histoire d'un génome: Population et génétique dans l'est du Québec* (Presses de l'Université du Québec, 1991).
38. J. Rondot, *Can. J. Earth Sci.* **5**, 1305–1317 (1968).
39. C. Moreau *et al*., *Science* **334**, 1148–1150 (2011).
40. D. Labuda, T. Harding, E. Milot, H. Vézina, *PLOS ONE* **17**, e0266079 (2022).
41. S. Schiffels, R. Durbin, *Nat. Genet.* **46**, 919–925 (2014).
42. E. M. L. Scerri, L. Chikhi, M. G. Thomas, *Nat. Ecol. Evol.* **3**, 1370–1372 (2019).
43. D. Nelson *et al*., *Am. J. Hum. Genet.* **103**, 893–906 (2018).
44. A. Gagnon, E. Heyer, *Am. J. Hum. Biol.* **13**, 645–659 (2001).
45. A. Gagnon, B. Toupance, M. Tremblay, J. Beise, E. Heyer, *Am. J. Phys. Anthropol.* **129**, 630–636 (2006).
46. K. M. Burkett *et al*., *Front. Genet.* **12**, 808829 (2022).
47. J. Wakeley, L. King, B. S. Low, S. Ramachandran, *Genetics* **190**, 1433–1445 (2012).
48. A. Bhaskar, A. G. Clark, Y. S. Song, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2385–2390 (2014).
49. H. Vézina, M. Jomphe, E.-M. Lavoie, C. Moreau, D. Labuda, *Cah. Que. Demogr.* **41**, 87–105 (2012).
50. L. Anderson-Trocmé, LukeAndersonTrocme/genes_in_space: v1.0, Zenodo (2023); https://doi.org/10.5281/zenodo.7703131.
51. L. Anderson-Trocmé, LukeAndersonTrocme/genome_simulations: v1.0, Zenodo (2023); https://doi.org/10.5281/zenodo.7705869.
52. Fichier BALSAC (Université du Québec à Chicoutimi), Genealogy of Quebec Population 1621-1993, version 1, Borealis: The Canadian Dataverse Repository (2022); https://doi.org/10.5683/SP3/BW7DIG.

ACKNOWLEDGMENTS

We are grateful to all the participants who enabled this study by contributing their DNA and to the participants who provided family information enabling the reconstruction of their genealogy. For data from Quebec, we thank the BALSAC team for its management and curation of the genealogy database and the CARTaGENE team and Genome Quebec for their management and curation of genotype data. For data from France, we thank the EREN team for its management and curation of SU.VI.MAX genotype data. We thank C. Bhérer, G. McVean, C. Moreau, A. Ragsdale, R. Sladek, W. Wohns, and Y. Wong for useful conversations. **Funding:** This work was supported in part by the Canadian Queen Elizabeth II Diamond Jubilee Scholarship (QES) (L.A.-T.) and the Fonds de recherche du Québec–Nature et technologies (FRQNT: B2X 290358) (L.A.-T.). This research was also supported by the Canadian Institute for Health Research (CIHR) project grant 437576, Natural Sciences and Engineering Research Council of Canada (NSERC) grant RGPIN-2017-04816, the Canada Research Chair program to S.G., and the Canada Foundation for Innovation. J.K. was supported by the Robertson Foundation. **Author contributions:** Conceptualization: L.A.-T., D.N., C.D., H.V., J.K., and S.G. Data curation: L.A.-T. and M.T. Formal analysis: L.A.-T. Funding acquisition: L.A.-T., J.K., and S.G. Investigation: L.A.-T. and S.G. Methodology: L.A.-T., D.N., S.Z., A.D.-P., J.K., and S.G. Project administration: L.A.-T., J.K., and S.G. Resources: S.G. Software: L.A.-T., D.N., I.K., S.Z., N.B., B.J., J.K., and S.G. Supervision: J.K., H.V., and S.G. Validation: L.A.-T., D.N., I.K., N.B., B.J., J.K., and S.G. Visualization: L.A.-T. Writing – original draft: L.A.-T. Writing – review & editing: L.A.-T., C.D., A.D.-P., S.Z., J.K., H.V., and S.G. **Competing interests:** The authors report no potential competing interests. **Data and materials availability:** R code for analyses and visualizations are available on Zenodo (https://zenodo.org/record/7703131) (*50*). Python code for running genome simulations are available on Zenodo (https://zenodo.org/record/7705869) (*51*). Tree sequences of simulated genomes are freely available on Zenodo (https://zenodo.org/record/7702392) (*34*). Quebec genotype data are available from the cohorts, via an independent data access committee by the CARTaGENE cohort (https://cartagene.qc.ca/en/researchers/access-request.html) and Genome Quebec for the Genizon cohort (https://genomequebec.com/0-genizon-biobank/). Genealogical data are available on the Scholars Portal Dataverse platform from the University of Québec in Chicoutimi (https://doi.org/10.5683/SP3/BW7DIG) (*52*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add5300
Materials and Methods
Supplementary Text
Figs. S1 to S24
Tables S1 to S4
References (*53*–*70*)
MDAR Reproducibility Checklist
French Translation of Author Accepted Manuscript

View/request a protocol for this paper from *Bio-protocol*.

Submitted 4 July 2022; accepted 24 April 2023
10.1126/science.add5300

BIOGEOGRAPHY

Erosion of heterogeneous rock drives diversification of Appalachian fishes

Maya F. Stokes[1,2,3]*, Daemin Kim[4], Sean F. Gallen[5], Edgar Benavides[4], Benjamin P. Keck[6], Julia Wood[4], Samuel L. Goldberg[2], Isaac J. Larsen[7], Jon Michael Mollish[8], Jeffrey W. Simmons[8], Thomas J. Near[4], J. Taylor Perron[2]

The high levels of biodiversity supported by mountains suggest a possible link between geologic processes and biological evolution. Freshwater biodiversity is high not only in tectonically active settings but also in tectonically quiescent montane regions such as the Appalachian Mountains. We show that erosion through different rock types drove allopatric divergence between lineages of the Greenfin Darter (*Nothonotus chlorobranchius*), a fish species endemic to rivers draining metamorphic rocks in the Tennessee River basin in the United States. In the past, metamorphic rock preferred by *N. chlorobranchius* was more widespread, but as erosion exposed other rock types, lineages of this species were progressively isolated in tributaries farther upstream, where metamorphic rock remained. Our results suggest a geologic mechanism for initiating allopatric diversification in mountains long after tectonic activity ceases.

Mountain ranges harbor a disproportionate amount of the world's biodiversity, suggesting that geologic processes may drive speciation (*1*–*3*). Topographic uplift during mountain building can isolate formerly contiguous populations, leading to allopatric speciation, and create novel habitats, generating opportunities for adaptive evolution. Uplift in tectonically active mountains such as the Andes (*1*, *4*) and Hengduan Mountains (*5*) has been linked to allopatric and adaptive speciation in terrestrial organisms (*5*–*9*). In freshwater fish, allopatric speciation has been linked to uplift in New Zealand (*7*) and the Pacific Rim of North America (*10*). However, older mountain ranges that have not experienced topographic uplift for hundreds of millions of years are also hotspots for freshwater biodiversity (*11*, *12*). The mechanisms driving diversification in these tectonically inactive landscapes are unclear. Here we investigate one such biodiverse and tectonically quiescent mountain range, the southern Appalachian Mountains of the United States.

The exceptional freshwater biodiversity of the southeastern United States (*11*, *13*) has been attributed to a variety of mechanisms, including low extinction rates due to environmental stability (*11*, *12*, *14*) and the presence of glacial refugia (*15*). The rearrangement of river networks may have driven allopatric speciation in fish (*16*, *17*) and salamanders (*18*). However, there is also evidence of diversification in fish lineages over small spatial scales in river basins with stable boundaries (*19*). In particular, the patterns of diversification within a darter fish lineage (*Nothonotus*) suggest that in situ speciation occurs in highland basins such as the upper Tennessee River (*20*). Yet it is unknown which within-basin geologic processes, if any, cause diversification of Appalachian fishes. In this study, we link the erosional exhumation of different rock types to the within-basin allopatric diversification of the Greenfin Darter (*Nothonotus chlorobranchius*), a species endemic to the upper Tennessee River.

The Tennessee River (Fig. 1) originates in the Blue Ridge geologic province, which is composed of high-grade metamorphic rock. Downstream, the river flows over the Valley and Ridge province, characterized by siliciclastic and carbonate sedimentary rock. The metamorphic rock of the Blue Ridge is harder to erode than the sedimentary rock of the Valley and Ridge, making the rivers in the Blue Ridge steeper (*21*). *N. chlorobranchius* is found almost exclusively in steeper rivers flowing over metamorphic rock within the Blue Ridge (Fig. 1 and fig. S1). However, the confluences between the Blue Ridge–draining tributaries are located farther downstream in the sedimentary rock of the Valley and Ridge, where *N. chlorobranchius* is absent. Thus, the populations located in different Blue Ridge tributaries are currently separated from each other by sections of the river network flowing over sedimentary rock. However, the spatial distribution of rocks at the surface of a landscape is not static. Erosion of nonvertical rock layers causes contacts between rock types to migrate across the landscape as the surface topography intersects those contacts in different locations (*22*, *23*).

In the Tennessee River, erosion has caused the exposure of sedimentary rock to expand, while the exposure of metamorphic rock has shrunk. Thus, in the past, suitable habitat for *N. chlorobranchius* may have been more widespread. Over time, erosion may have gradually isolated populations in different tributaries where metamorphic rock persisted, catalyzing allopatric lineage diversification. To test this hypothesis, we analyzed phylogenomic data to determine whether populations of *N. chlorobranchius* in Blue Ridge tributaries are distinct evolutionary lineages. We then compared the evolutionary history of diversification to the geologic history of the upper Tennessee River basin using a geometric model of bedrock erosion.

Phylogeography of *N. chlorobranchius*

We inferred a phylogeny of the darter clade *Nothonotus* using DNA sequence data sampled from across the genome acquired with a double-digest restriction site–associated DNA (ddRAD) methodology (*24*). The results confirm that *N. chlorobranchius* is monophyletic (all individuals form a single and unique clade) and that its sister lineage is a clade composed of the Bluebreast Darter (*N. camurus*) and the Orangefin Darter (*N. bellus*) (fig. S2). We investigated the phylogeography of lineages within *N. chlorobranchius* using both ddRAD sequence data and a mitochondrial gene (mtDNA).

The phylogenomic analyses support the interpretation that *N. chlorobranchius* populations in different tributaries of the upper Tennessee River are isolated from one another. Both the ddRAD and mtDNA phylogenies resolve most *N. chlorobranchius* individuals sampled from a given Blue Ridge tributary as monophyletic (Fig. 2). Additional population genetic analyses of the ddRAD dataset also indicate a lack of gene flow between most tributaries. Between-tributary values of the fixation index (F_{ST}), a measure of genetic differentiation, are higher (median = 0.6) than within-tributary values (median = 0.04) according to a Wilcoxon rank sum test ($P < 0.001$), a pattern consistent with a history of isolation between tributaries (fig. S3A). Similarly, genetic divergence (D_{CH}) is positively correlated with streamwise distance (D_{ST}) when comparing across different tributaries but not when comparing within tributaries (fig. S3B), supporting the hypothesis of a between-tributary barrier.

We used two species-delimitation methods to assess whether the lineages of *N. chlorobranchius* exhibit the divergence expected among distinct species. Values of the genealogical divergence index are intermediate (0.3 to 0.7), consistent with either population-level or species-level divergence (fig. S3C) (*25*). An

[1]Yale Institute for Biospheric Studies, New Haven, CT 06511, USA. [2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL 32304, USA. [4]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511, USA. [5]Department of Geosciences, Colorado State University, Fort Collins, CO 80523, USA. [6]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA. [7]Department of Earth, Geographic, and Climate Sciences, University of Massachusetts, Amherst, MA 01003, USA. [8]Fisheries and Aquatic Monitoring, Tennessee Valley Authority, Chattanooga, TN 37415, USA.
*Corresponding author. Email: mfstokes@fsu.edu

assessment of morphological traits (*24*) shows that the lineages are not morphologically distinguishable (fig. S6). Taken together, the similar geologic setting, high levels of genetic divergence, and morphological similarity provide support for an allopatric diversification mechanism among the geographically isolated *N. chlorobranchius* lineages (*26*).

Three exceptions to these patterns suggest a more nuanced evolutionary history of *N. chlorobranchius* in some tributaries. First, there are two distinct clades resolved among

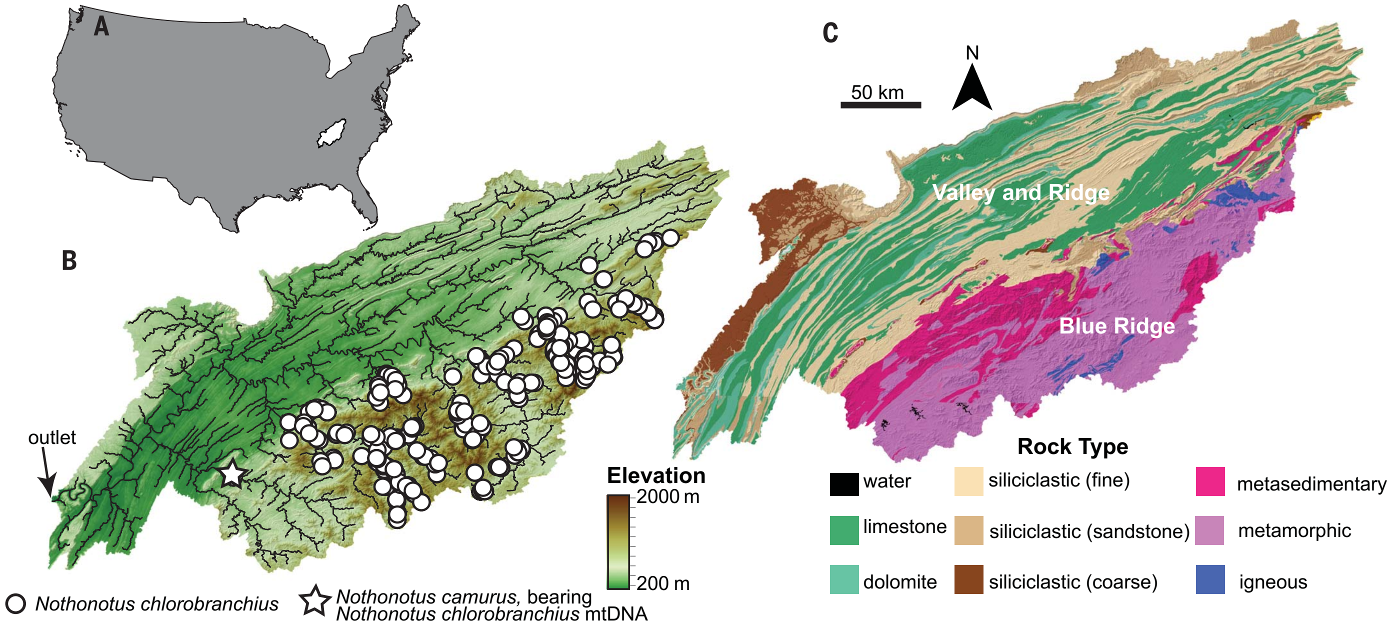


Fig. 1. Map of study area. (**A**) Location of the upper Tennessee River basin (white) within the conterminous United States. (**B**) Topographic map of the basin showing the river network (black lines) and localities where *N. chlorobranchius* has been observed (circles). The star shows the locality of a specimen of *N. camurus* bearing a *N. chlorobranchius*–like mitochondrial gene. (**C**) Geologic map of the basin.

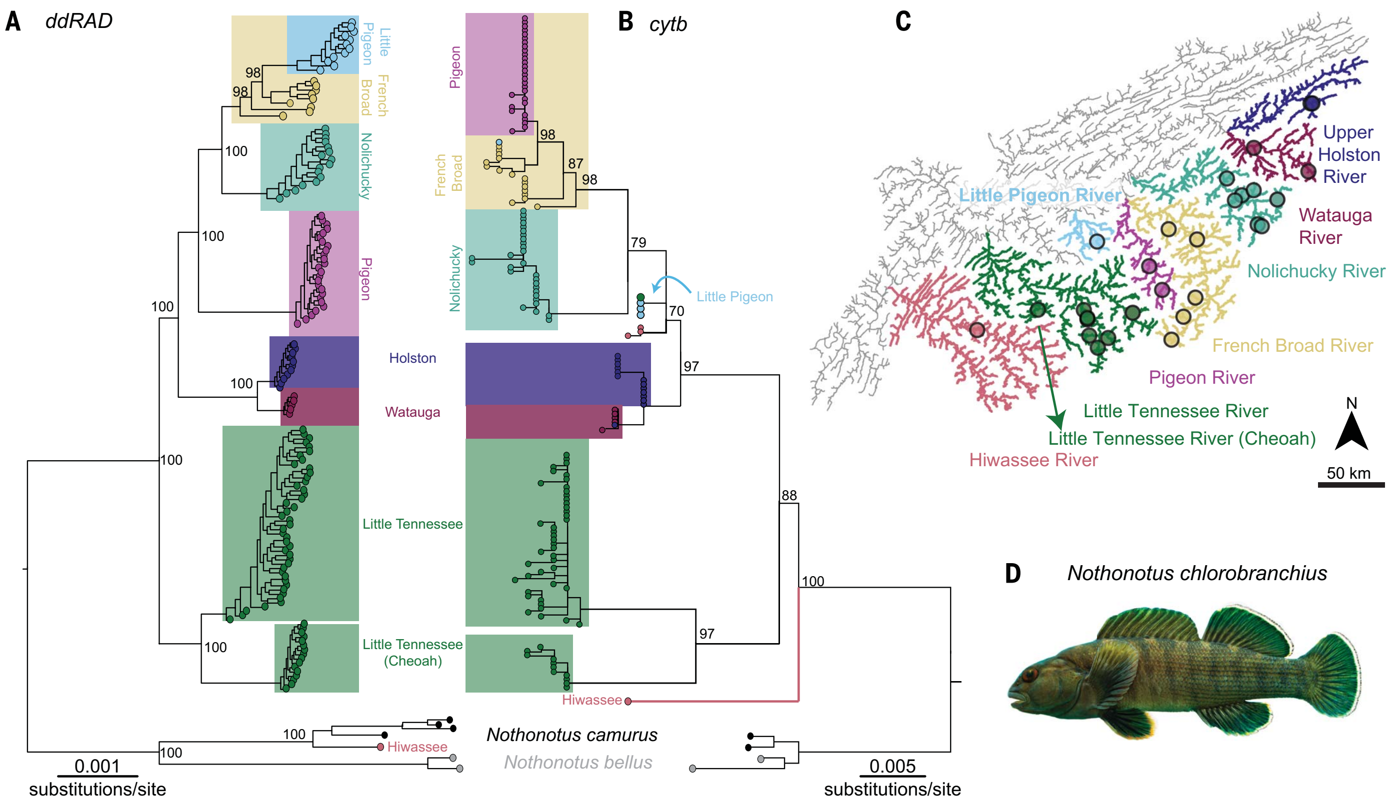


Fig. 2. Phylogeography of *N. chlorobranchius*. Phylogenies inferred from (**A**) concatenated ddRAD sequence data and (**B**) the mtDNA gene *cytb*. The tips of the trees are colored according to the tributary where each specimen was sampled. Clades are outlined with rectangles colored to correspond with the tributaries in (**C**). Nodes are labeled with bootstrap support values. (**D**) A male specimen from the Pigeon River [Photo: Jon Michael Mollish].

specimens from the Little Tennessee River, one in the mainstem and the other in a tributary, the Cheoah River. These two clades exhibit the positive D_{CH} versus D_{ST} relationship and high F_{ST} value (F_{ST} = 0.6) observed for between-tributary comparisons (fig. S3, A and B), yet there is no sedimentary rock exposed between them, suggesting that streamwise distance is an isolating mechanism within the Little Tennessee River.

Second, there is evidence for gene flow between some of the tributaries. For the French Broad and Little Pigeon rivers, F_{ST} = 0.2, which is lower than for other between-tributary comparisons (fig. S3A). Gene flow between these two tributaries is also evident in the phylogenies. In the ddRAD phylogeny, individuals from the Little Pigeon River are nested within the French Broad clade (Fig. 2A), and in the mtDNA tree, Little Pigeon River haplotypes (mtDNA variants) are mostly placed within a Nolichucky River clade (Fig. 2B). We conclude that there has been a history of dispersal to the Little Pigeon from neighboring basins. Similarly, Pigeon River mtDNA haplotypes are nested within a French Broad River clade (Fig. 2B), indicating a history of migration consistent with topographic evidence of a past river capture between these two basins (*21*).

Third, there is an important difference between the two phylogenies in the resolution of a specimen sampled from the southernmost Blue Ridge tributary, the Hiwassee River. *N. chlorobranchius* is not present in the Hiwassee, but a closely related species, *N. camurus,* is (*27*). The two species have neighboring but typically nonoverlapping distributions, with *N. chlorobranchius* found in rivers draining the Blue Ridge and *N. camurus* in the Valley and Ridge (fig. S4). Where their ranges overlap, previous work indicates hybridization (*28*). We obtained only one individual from the Hiwassee for our genetic analyses, which was identified as *N. camurus* in the field. In the ddRAD phylogeny, this specimen exhibits a genomic ancestry consistent with other sampled *N. camurus*. In contrast, the mtDNA haplotype sequenced from this specimen exhibits ancestry consistent with *N. chlorobranchius* (Fig. 2B). This same *N. chlorobranchius*–like mtDNA haplotype is also observed in specimens of a more distantly related species, the Redline Darter (*N. rufilineatus*), collected from the Hiwassee (fig. S5). Incongruence between phylogenies inferred from mitochondrial and nuclear genomes can arise as a result of hybridization (*29*–*32*). We therefore interpret the Hiwassee *N. camurus* specimen as evidence of an ancient and now absent Hiwassee River lineage of *N. chlorobranchius* that hybridized with *N. rufilineatus* and *N. camurus* while they coexisted.

Erosion through different rock types and the restriction of dispersal pathways

The geology and topography of the upper Tennessee River basin provide an additional test of the hypothesized link between geology and biological evolution. If the populations of *N. chlorobranchius* in different tributaries are currently isolated by barriers of exposed sedimentary rock, how did their ancestral population disperse throughout the river network? The Blue Ridge metamorphic rocks are located within thrust sheets dipping toward the southeast (SE) and overlying the sedimentary rocks of the Valley and Ridge (*33*, *34*). As surface erosion exhumed the contact between sedimentary and metamorphic rock, the contact migrated upstream (Fig. 3A). Although the confluences between Blue Ridge tributaries are currently located in sedimentary rock, they were likely located in metamorphic rock in the past, allowing for dispersal between populations now isolated in the upper reaches of these rivers (Fig. 3B). We posit that this process

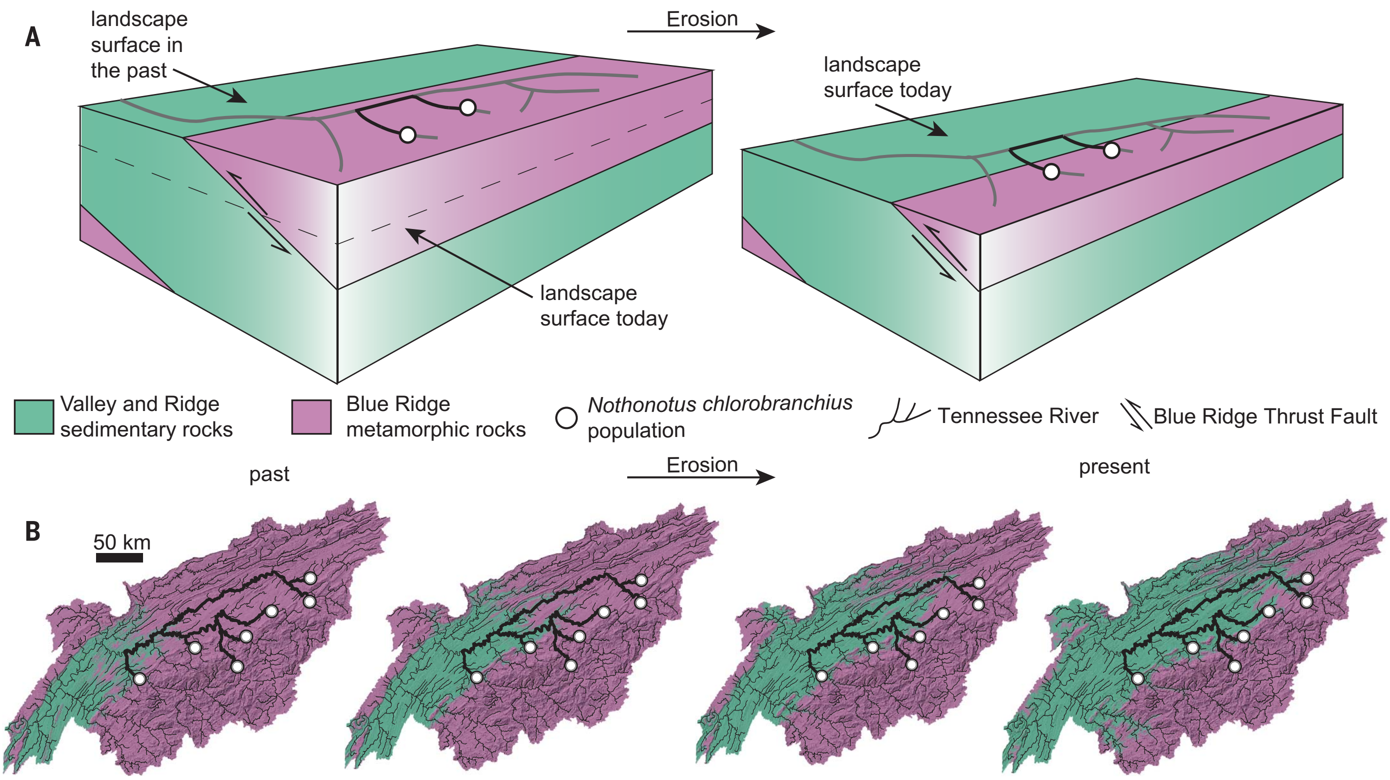


Fig. 3. Migration of a rock contact as a result of erosion. (**A**) Conceptual block diagrams showing a simplified cross-sectional view of the geologic structure of the upper Tennessee River basin in the past (left) and the present (right). As the rock above the dashed line erodes away, the contact between sedimentary and metamorphic rocks migrates across the landscape, such that a streamwise path between two hypothetical populations (bold line) is initially located entirely in metamorphic rocks and later crosses sedimentary rocks. (**B**) Reconstruction of rock contact migration, overlaid on a shaded relief map of the basin. Bold lines show streamwise paths between the most downstream instance of *N. chlorobranchius* recorded in each of the studied tributaries. See fig. S7 and (*24*) for more details.

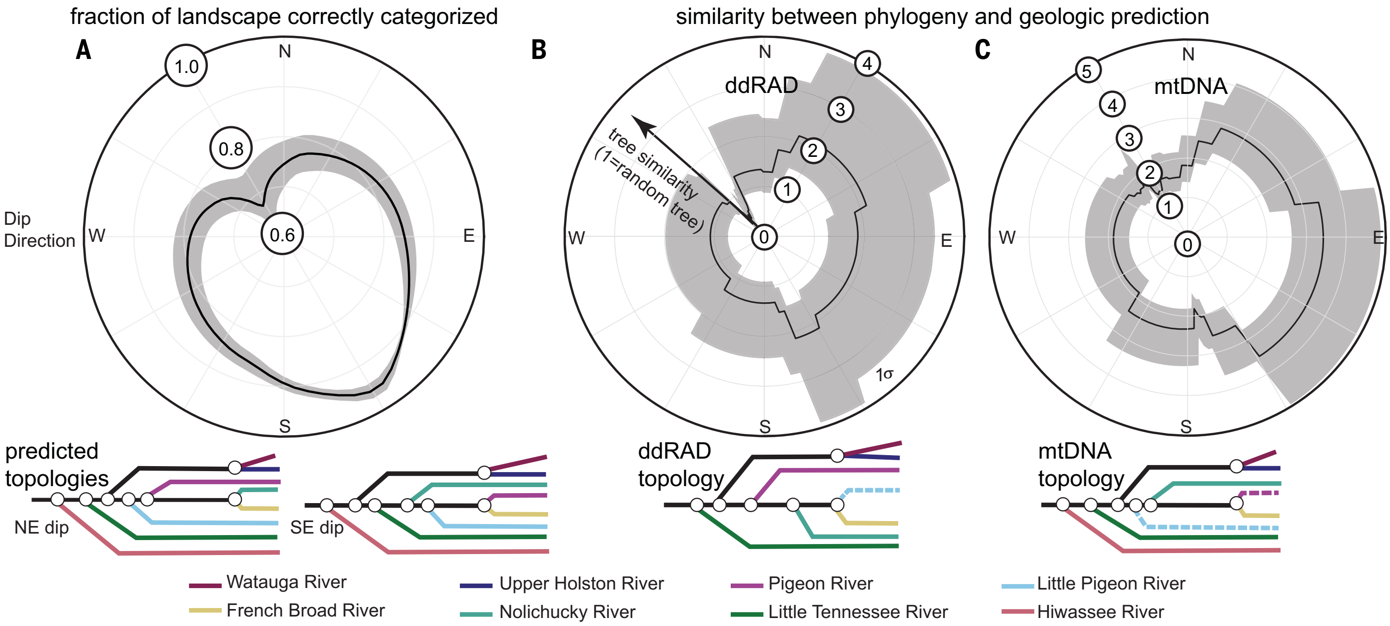


Fig. 4. Comparison of phylogenies to geologic evolutionary scenarios. (**A**) The fraction of the landscape for which a geologic contact with a given dip direction correctly predicts the rock type (metamorphic or sedimentary) exposed at the surface. (**B** and **C**) Mutual clustering information between dendrograms representing the topologic relationships of lineages inferred from the ddRAD phylogeny (B) and the mtDNA phylogeny (C) and dendrograms predicted by erosion into a geologic contact, normalized by similarity to 1000 randomly generated dendrograms. Dashed lines in the dendrograms in (B) and (C) indicate tributaries where specimens were not resolved as monophyletic. In all polar plots, the black line represents the mean, and the shaded area represents ±1σ.

drove geographic isolation between tributaries and was the mechanism of allopatric diversification among *N. chlorobranchius* lineages.

Comparison of biological and geological evidence

To test this hypothesis, we compared the topology of the mtDNA and ddRAD phylogenies (Fig. 2) to dendrograms representing an evolutionary scenario produced by a simplified model of the hypothesized geological mechanism: erosion into a dipping, planar contact between sedimentary and metamorphic rock. To construct the geologic dendrograms, we assumed a steady and uniform erosion rate and estimated the relative timing of exposure of sedimentary rocks along streamwise paths between the tributaries (fig. S7) (*24*). The geologic model requires knowledge of the direction and magnitude of the dip of the planar contact between metamorphic and sedimentary rock. We find that a plane dipping toward the SE best predicts the present-day exposure of metamorphic rock (Fig. 4A), a result consistent with seismic reflection data indicating a shallow SE dip, although local heterogeneities exist (*35*).

A dip direction toward the northeast (NE) or SE produces dendrograms that are topologically similar to the ddRAD phylogeny (Fig. 4B), whereas SE and east (E) dip directions best match the mtDNA phylogeny (Fig. 4C) (*24*). Excluding nonmonophyletic tributaries produces largely similar results but favors a NE dip over a SE dip for the ddRAD phylogeny (fig. S8). The Hiwassee River specimen is excluded from the analysis of the ddRAD phylogeny because of its resolution as *N. camurus* (Fig. 2). However, because this specimen contains *N. chlorobranchius*–like mtDNA, it is included in analysis of the mtDNA tree, increasing the degree of similarity to the geologic model (Fig. 4). In all analyses, different dip directions produce dendrograms matching different features of the phylogenetic trees. A SE dip better resolves the relationships between the Pigeon, Nolichucky, and French Broad rivers observed in the ddRAD phylogeny; however, an E or NE dip better resolves the relationship between the Holston, Watauga, and Little Tennessee rivers observed in both phylogenies (Fig. 4). Spatial variability in the attitude of the geologic structure across the basin could explain why both NE and E or SE dip directions are consistent with different aspects of the phylogenies. Despite these nuances, the geologically realistic SE dip produces an evolutionary scenario consistent with both phylogenies, supporting the hypothesis that migration of the geologic contact was the mechanism of allopatric divergence among *N. chlorobranchius* lineages.

We additionally compared molecular and geologic estimates for the age of the most recent common ancestor of all *N. chlorobranchius*, which corresponds to the time when the Little Tennessee River was isolated from the other tributaries (Fig. 2). A relaxed molecular clock age approach using concatenated ddRAD loci produces an age estimate of 2.99 to 5.36 million years ago (Ma), and a multispecies coalescent model estimates a slightly younger age range of 2.38 to 4.07 Ma (figs. S10 and S11) (*24*). The geologic estimate for the timing of isolation of the Little Tennessee River is a function of how quickly the landscape is eroding and how steeply the contact between metamorphic and sedimentary rock is dipping (fig. S11). The geologic and molecular estimates for the absolute timing of isolation overlap if the dip angle is low (<0.3° relative to horizontal), and the erosion rates are between 43 and 70 m per million years. A least-squares fit of a plane to the surface trace of the contact between rock types supports a similarly shallow dip angle (~0.2° relative to horizontal). Erosion rates in the region average 18 m per million years, slower than required for the overlap between the geologic and molecular timing estimates (fig. S11). However, there are several uncertainties we do not account for in our analyses. We assume a constant and uniform erosion rate, but some evidence suggests that erosion was faster during glacial periods (*36*, *37*). Similarly, perturbations in base-level elevation could have produced pulses of faster river erosion (*21*). Additionally, the fossil record of darters does not provide any calibration points for molecular clock analyses; instead, divergence time

analysis relies on calibrations from the fossil-rich freshwater fish lineage Centrarchidae, which diverged from darters more than 30 Ma (*19*, *30*). Despite these challenges, the approximate overlap between these independent estimates in absolute time, in conjunction with the ability of the model of bedrock erosion to predict an evolutionary scenario with a topology consistent with phylogenetic relationships derived from molecular datasets (Fig. 4), provides support for rock type–mediated diversification of *N. chlorobranchius*.

Mechanisms of isolation and diversification

Our approach shows that erosion into layers of rock with different properties can drive allopatric lineage divergence in freshwater fish within a single river basin, revealing a mechanism that could drive the diversification of freshwater species in other tectonically inactive mountains. The specific effects of rock type on freshwater fish may vary depending on the species involved and the geologic context. For example, in New Zealand, the presence of gorges in schist has been linked to lower within-basin population connectivity (*38*). In the Tennessee River, characteristics of the flow and channel bed are likely important for species of *Nothonotus*, which tend to live in fast-flowing riffles (*27*). Indeed, channels flowing over metamorphic rock, where *N. chlorobranchius* is found (Fig. 1), are steeper, shallower, smaller in cross section, and likely have coarser bed sediment than comparably sized rivers flowing over sedimentary rock (fig. S1). Even so, the exact link between rock type and the ecology of *N. chlorobranchius* is unknown. Habitat specificity can arise from adaptive or nonadaptive processes (*39*), and different species of *Nothonotus* have different relationships to geologic features. For example, the closely related species *N. camurus* is generally restricted to sedimentary rock in the Tennessee River (fig. S4). It is possible that the range of *N. camurus* has expanded along with the exposure of sedimentary rock, and its presence could be a biotic barrier to *N. chlorobranchius* that mirrors the abiotic patterns described in this study. Regardless of the mechanism, specialization to specific rock types provided the precondition for the ongoing allopatric diversification observed among *N. chlorobranchius* lineages. The case of *Nothonotus* in the Tennessee River highlights the opportunity to integrate ecological knowledge with genomic and geologic datasets to fully decipher the impact of geologic processes on the diversification of freshwater species.

REFERENCES AND NOTES

1. C. Hoorn, V. Mosbrugger, A. Mulch, A. Antonelli, *Nat. Geosci.* **6**, 154–154 (2013).
2. A. Antonelli *et al.*, *Nat. Geosci.* **11**, 718–725 (2018).
3. C. Rahbek *et al.*, *Science* **365**, 1108–1113 (2019).
4. C. Hoorn *et al.*, *Science* **330**, 927–931 (2010).
5. W. N. Ding, R. H. Ree, R. A. Spicer, Y. W. Xing, *Science* **369**, 578–581 (2020).
6. J. T. Weir, M. Price, *Mol. Ecol.* **20**, 4550–4563 (2011).
7. D. Craw, P. Upton, C. P. Burridge, G. P. Wallis, J. M. Waters, *Nat. Geosci.* **9**, 140–144 (2015).
8. L. P. Lagomarsino, F. L. Condamine, A. Antonelli, A. Mulch, C. C. Davis, *New Phytol.* **210**, 1430–1442 (2016).
9. L. M. Boschman, F. L. Condamine, *Global Planet. Change* **208**, 103704 (2022).
10. D. R. Montgomery, *Geology* **28**, 1107–1110 (2000).
11. J. G. Lundberg, M. Kottelat, G. R. Smith, M. L. J. Stiassny, A. C. Gill, *Ann. Mo. Bot. Gard.* **87**, 26–62 (2000).
12. D. Griffiths, *J. Biogeogr.* **45**, 784–792 (2018).
13. J. April, R. L. Mayden, R. H. Hanner, L. Bernatchez, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10602–10607 (2011).
14. G. R. Smith, C. Badgley, T. P. Eiting, P. S. Larson, *Evol. Ecol. Res.* **12**, 693–726 (2010).
15. B. M. Burr, L. M. Page, in *The Zoogeography of North American Freshwater Fishes*, C. H. Hocutt, E. O. Wiley, Eds. (Wiley, 1986).
16. R. L. Mayden, *Syst. Zool.* **37**, 329–355 (1988).
17. C. M. Bossu, J. M. Beaulieu, P. A. Ceas, T. J. Near, *Mol. Ecol.* **22**, 5397–5417 (2013).
18. K. H. Kozak, R. A. Blaine, A. Larson, *Mol. Ecol.* **15**, 191–207 (2006).
19. P. R. Hollingsworth Jr., T. J. Near, *Evolution* **63**, 228–243 (2009).
20. T. J. Near, B. P. Keck, *Mol. Ecol.* **14**, 3485–3496 (2005).
21. S. F. Gallen, *Earth Planet. Sci. Lett.* **493**, 150–160 (2018).
22. M. Perne, M. D. Covington, E. A. Thaler, J. M. Myre, *Earth Surf. Dyn.* **5**, 85–100 (2017).
23. A. M. Forte, B. J. Yanites, K. X. Whipple, *Earth Surf. Process. Landf.* **41**, 1736–1757 (2016).
24. Materials and methods are available as supplementary materials.
25. N. D. Jackson, B. C. Carstens, A. E. Morales, B. C. O'Meara, *Syst. Biol.* **66**, 799–812 (2017).
26. S. A. S. Anderson, J. T. Weir, *Science* **378**, 1214–1218 (2022).
27. D. A. Etnier, W. C. Starnes, *The Fishes of Tennessee* (Univ. of Tennessee Press, 1993).
28. D. J. Eisenhour, *Copeia* **1995**, 368–379 (1995).
29. B. P. Keck, T. J. Near, *Evolution* **64**, 1410–1428 (2010).
30. T. J. Near *et al.*, *Syst. Biol.* **60**, 565–595 (2011).
31. T. J. Near, B. P. Keck, *Mol. Phylogenet. Evol.* **66**, 868–876 (2013).
32. R. C. Harrington, J. W. Simmons, T. J. Near, *Bull. Peabody Mus. Nat. Hist.* **61**, 3–21 (2020).
33. F. A. Cook *et al.*, *Geology* **7**, 563–567 (1979).
34. F. A. Cook, K. Vasudevan, *Tectonophysics* **420**, 161–174 (2006).
35. R. D. Hatcher, P. J. Lemiszki, J. B. Whisner, *Spec. Pap. Geol. Soc. Am.* **433**, 243–276 (2007).
36. P. Molnar, *Annu. Rev. Earth Planet. Sci.* **32**, 67–89 (2004).
37. D. M. Anthony, D. E. Granger, *Earth Surf. Process. Landf.* **32**, 874–887 (2007).
38. J. M. Waters *et al.*, *J. Biogeogr.* **42**, 2452–2460 (2015).
39. J. M. Waters, B. C. Emerson, P. Arribas, G. A. McCulloch, *Trends Ecol. Evol.* **35**, 512–522 (2020).

ACKNOWLEDGMENTS

We thank G. Watkins-Colwell for assistance with research collections at the Yale Peabody Museum of Natural History and three anonymous reviewers for suggestions that helped us improve the manuscript. **Funding:** J.T.P. and M.F.S. acknowledge the support of the MIT mTerra Catalyst Fund and the National Science Foundation through the EarthScope AGeS Program and the Graduate Research Fellowship Program under grant 1122374. S.F.G. acknowledges support from Colorado State University, the Warner College of Natural Resources, and the Department of Geosciences. T.J.N. acknowledges support from the Bingham Oceanographic Endowment maintained by the Peabody Museum, Yale University. **Author contributions:** M.F.S., J.T.P., T.J.N., and S.F.G. designed the research. M.F.S., T.J.N., D.K., B.P.K., J.M.M., and J.W.S. collected fish specimens. J.W.S. contributed all Tennessee Valley Authority presence and absence data. M.F.S., E.B., B.P.K., and D.K. collected DNA sequence data. J.W. and B.P.K. collected morphological data. M.F.S., S.F.G., and J.T.P. collected sediment samples. M.F.S. and I.J.L. conducted lab work to measure cosmogenic nuclide–derived erosion rates. S.L.G. and M.F.S. compiled a database of published erosion rates. M.F.S. conducted formal analysis of geologic and genomic data. M.F.S., D.K., and T.J.N. conducted formal analysis of genomic data. M.F.S. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The raw ddRAD sequence data are available at the NCBI BioProject Archive (BioProjectID PRJNA910655) and *cytb* data are available on GenBank. All other data are available in the supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add9791
Materials and Methods
Figs. S1 to S12
Tables S1 to S6
References (*40*–*89*)
MDAR Reproducibility Checklist
Data S1 to S4

Submitted 17 July 2022; accepted 25 April 2023
10.1126/science.add9791

2D MATERIALS

Correlated insulator of excitons in WSe_2/WS_2 moiré superlattices

Richen Xiong[1], Jacob H. Nie[1], Samuel L. Brantly[1], Patrick Hays[2], Renee Sailus[2], Kenji Watanabe[3], Takashi Taniguchi[4], Sefaattin Tongay[2], Chenhao Jin[1]*

A panoply of unconventional electronic states has been observed in moiré superlattices. Engineering similar bosonic phases remains, however, largely unexplored. We report the observation of a bosonic correlated insulator in tungsten diselenide/tungsten disulfide (WSe_2/WS_2) moiré superlattices composed of excitons, that is, tightly bound electron-hole pairs. We develop a pump probe spectroscopy method that we use to observe an exciton incompressible state at exciton filling $\nu_{ex} = 1$ and charge neutrality, indicating a correlated insulator of excitons. With varying charge density, the bosonic correlated insulator continuously transitions into an electron correlated insulator at charge filling $\nu_e = 1$, suggesting a mixed correlated insulating state between the two limits. Our studies establish semiconducting moiré superlattices as an intriguing platform for engineering bosonic phases.

Strongly correlated phases can emerge in flat-band systems when many-body interactions dominate over kinetic energy (*1*, *2*). Semiconducting transition-metal dichalcogenide (TMDC) moiré superlattices offer a distinctive platform where both fermionic and bosonic quasiparticles—charges (*3–12*) and excitons (*13–21*)—occupy flat bands. Previous studies have primarily focused on the fermionic sector with exciton filling $\nu_{ex} = 0$, such as Mott and Wigner crystal states (*3–8*), stripe phase (*9*), continuous Mott transition (*10*, *11*), and quantum anomalous Hall insulators (*12*), along with single-exciton behavior in the $\nu_{ex}\rightarrow 0$ limit, such as moiré excitons (*13–16*) and trions (*20*, *21*). Recently, increasing efforts have been put into an intermediate exciton density regime of $\nu_{ex} \sim 0.1$. Examples include studies of exciton-mediated ferromagnetism (*22*) as well as excitonic insulators (*23*, *24*)—insulators for charge but "metals" for excitons, where electron behavior is affected by the coexisting excitons. However, in these prior studies, excitons themselves are in a compressible fluid state; strongly correlated phases of bosons remain elusive.

Here we explore the high–exciton density regime of $\nu_{ex} \sim 1$. We create and identify a bosonic correlated insulator consisting of interlayer excitons in a 60°-aligned WSe_2/WS_2 moiré superlattice. Each interlayer exciton contains an electron in the WS_2 layer and a hole in WSe_2 with a large binding energy of hundreds of milli–electron volts (*25*), which is the ground state exciton configuration of a type II heterostructure (Fig. 1A). We find an exciton incompressible state at $\nu_{ex} = 1$, i.e., one exciton per moiré site, a hallmark of a bosonic correlated insulator. We further study the phase diagram spanned by ν_e and ν_{ex} and observe a mixed correlated insulator along the path of $\nu_{tot} = \nu_{ex} + \nu_e = 1$. Owing to the large exciton binding energy and strong correlations in this system, the bosonic correlated insulator persists to above 30 K, orders of magnitude higher than in previous studies in cold atoms and quantum wells (*26*, *27*). Our results highlight semiconducting moiré superlattices as an attractive platform for engineering new bosonic phases at high temperature, such as valley pseudospin order and pseudospin liquid, bosonic Mott-superfluidity transition (*28*), exciton-mediated superconductivity (*29*), and topological excitons (*18*, *19*), as well as for exploring the many-body physics of interacting fermions and bosons.

[1]Department of Physics, University of California at Santa Barbara, Santa Barbara, CA 93106, USA. [2]School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ 85287, USA. [3]Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [4]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan.
*Corresponding author. Email: jinchenhao@ucsb.edu

Direct measurement of exciton compressibility

One major challenge in identifying a correlated insulator of excitons is to distinguish it from disordered excitons without a lattice. Several works have reported power-dependent photoluminescence (PL) spectra in TMDC systems (*30*, *31*) and emergence of additional high-energy peaks under strong excitation. However, this only indicates the existence of local exciton-exciton interactions while providing no information on an ordered exciton lattice. The key evidence of a correlated insulator, as widely recognized in electrical measurements, is an incompressible state at particular lattice fillings (*32*). Optical measurements such as PL, by contrast, typically collect responses from all excitons in the system and cannot obtain compressibility information. Here we develop a pump-probe spectroscopy method that directly measures exciton compressibility (Fig. 1B); this is an optical analog of electrical capacitance measurements (*33*) and distinct from conventional optical pump-probe configurations. In an electronic capacitance measurement, a DC "pump" gate voltage tunes the background charge density, and a small AC "probe" voltage slightly modulates the charge density. Similarly, here a relatively strong pump light tunes the background exciton density, and the weak probe light injects a small number of additional excitons and detects their response. In both cases, the charge and exciton compressibility can be directly obtained from the minimum energy it takes to add one more particle on top of a given background particle density. Notably, a DC "pump" is necessary to maintain a stable background particle density and a well-defined ground state, whereas the "probe" needs to be AC modulated to isolate the responses of particles created by the probe (*34*).

Correlated insulator of bosons

With the capability of tuning charge and exciton density through electrostatic gating and pump light, respectively, we fully explore the phase diagram spanned by the charge filling ν_e and exciton filling ν_{ex}. We start with the axes, i.e., $\nu_{ex} = 0$ or $\nu_e = 0$. Figure 1, C and E, show the PL and absorption spectra, respectively, of a 60°-aligned moiré bilayer (device D1) at $\nu_{ex} = 0$ (zero pump intensity) and $\nu_e \geq 0$ (see fig. S2 for complete doping dependence). At charge neutrality, the PL features a single peak at 1.43 eV from interlayer exciton emission (*6*, *8*), whereas the absorption shows three peaks from moiré intralayer excitons (*13*). At approximately ν_e = 1 and 2 [$\nu_e = 1$ corresponds to the moiré density $n_0 = 2.1 \times 10^{12}$ cm^{-2} (*34*)], the emission peak blueshifts suddenly, and the absorption peaks show a kink. These features originate from the emergence of insulating states at integer electron fillings, which has been independently confirmed by capacitance and microwave impedance measurements (*3*, *6*). The PL energy jump at $\nu_e = 1$ can be intuitively understood from the emergence of a fermionic correlated insulator state in which all available sites are occupied by one electron; any additional excitons injected are therefore forced into a higher-energy state. We will discuss the mechanisms of these spectral changes in more depth in the discussion section.

We now turn to the case of $\nu_e = 0$ by fixing the gate voltage V_g at charge-neutral −0.5 V. Figure 1, D and F, show the dependence of pump-probe PL and absorption spectra on the pump light intensity, which effectively controls ν_{ex}. To account for the nonlinear dependence of ν_{ex} on pump intensity, we also show on the right axes dipolar-interaction–induced interlayer exciton energy shift Δ_{dipole} (the energy shift of peak I relative to zero pump intensity;

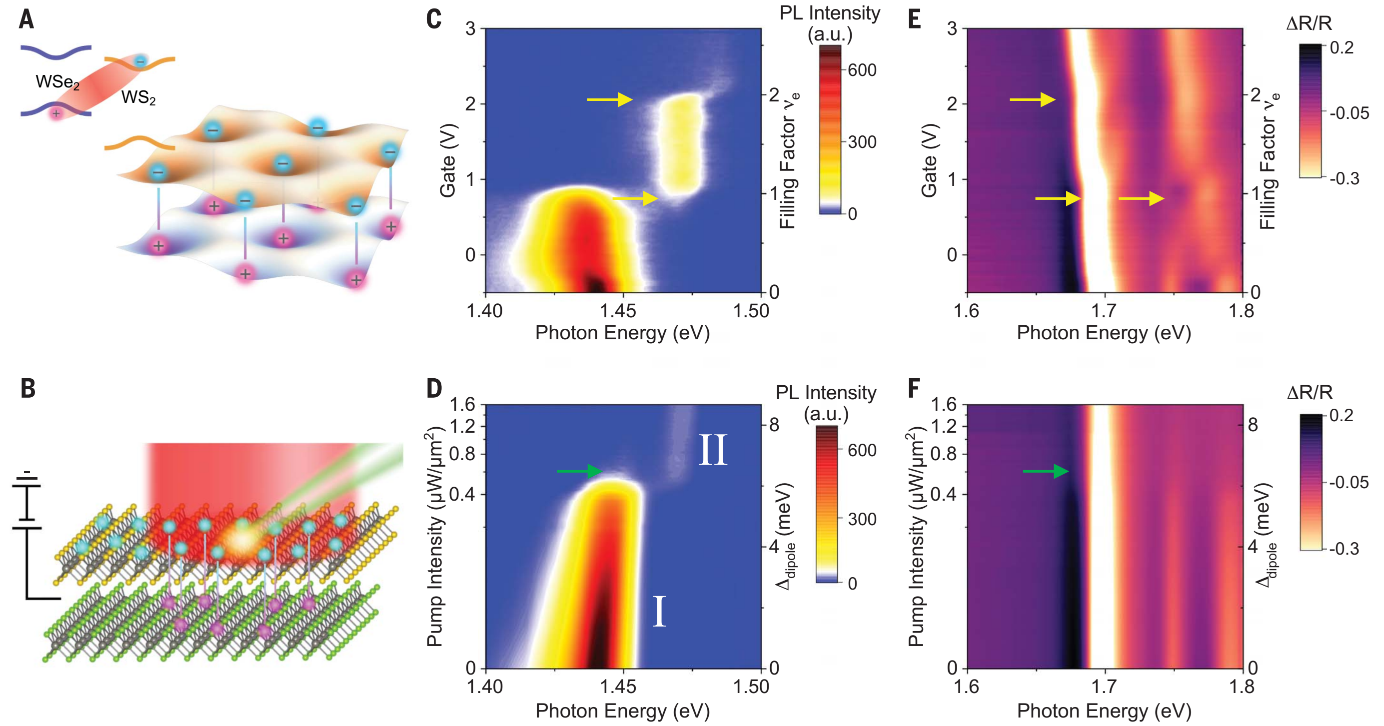


Fig. 1. Bosonic correlated insulator. (**A**) Illustration of a bosonic correlated insulator consisting of interlayer excitons. Magenta spheres indicate holes and cyan spheres, electrons. (Inset) Type II band alignment of WSe_2/WS_2 heterostructure. (**B**) Schematics of continuous-wave pump probe spectroscopy. The exciton and electron density are independently controlled by pump light and electrostatic gate. Red and green shading correspond to wide-field pump light and focused probe light, respectively. (**C** and **E**) Gate-dependent PL (C) and absorption (E) spectra of a 60°-aligned WSe_2/WS_2 moiré bilayer (device D1) at zero pump intensity. The PL peak shows a sudden blue shift at electron filling ν_e= 1 and 2 (yellow arrows), where the absorption spectrum shows kinks and splitting. (**D** and **F**) Pump intensity–dependent PL (D) and absorption (F) spectra of device D1 at charge neutrality. Right axes show dipolar-interaction–induced interlayer exciton energy shift Δ_{dipole}, which is approximately proportional to ν_{ex}. The dominant PL peak in (D) at low and high pump intensity are labeled as peak I and II, respectively. All measurements are performed at a base temperature of 1.65 K.

see Fig. 1D), which is approximately proportional to ν_{ex} (*34–36*). Notably, a jump in the PL energy is observed here as well. This jump is well reproduced in two other 60°-aligned moiré bilayers and one 0°-aligned moiré bilayer but is absent in a slightly misaligned bilayer (*34*), indicating that its origin is from correlation effects. The qualitative similarity to the gate-dependent PL suggests a similar origin behind the exciton energy jump: the emergence of a particle lattice that occupies all available sites. The most natural candidates are interlayer excitons, which are the immediate products of pump light absorption, and interlayer charge transfer in a type II heterojunction. To elucidate the nature of the lattice, we compare its effects on PL and absorption spectra to those from an electron lattice. PL spectra probe interlayer exciton responses. Although both the gate-induced electron lattice (Fig. 1C) and pump-induced lattice (Fig. 1D) lead to a jump in interlayer exciton energy in PL, the amplitudes are quite different: 35 meV and 15 meV, respectively. A more prominent difference is observed in the absorption spectra, which examine how intralayer excitons respond to the induced lattices. Electrons induce rich features, such as shifting, merging, and splitting of exciton resonances (Fig. 1E). By contrast, the pump-injected particles have rather weak effects on intralayer excitons, only slightly decreasing their oscillator strength (Fig. 1F). These distinctive behaviors indicate that the pump-induced lattice is not formed by electrons.

To further confirm the exciton nature of the pump-induced lattice, we move away from the axes and investigate the phase diagram at ν_{ex} > 0 and ν_e > 0. Figure 2A shows a set of pump intensity-dependent PL spectra at different gate voltages, corresponding to increasing electron density in the WS_2 layer. All plots have qualitatively similar behaviors—interlayer exciton energy shows a jump. Upon increasing ν_e, the jump occurs earlier and earlier, eventually appearing at zero pump intensity when $\nu_e = 1$ (V_g = 0.8 V), consistent with the gate-injected electrons having already occupied all sites at $\nu_e = 1$. On the basis of these observations, the pump-injected particles will occupy sites previously available to free electrons in WS_2 and form a mixed lattice together with free electrons. Besides an electron itself, which has been already excluded, the only other candidate for the pump-injected particle is the interlayer exciton composed of an electron in the WS_2 layer and a hole in the WSe_2 layer.

We, therefore, conclude that the pump light is creating interlayer excitons and tuning ν_{ex}. At charge neutrality, the observed jump in PL spectra then corresponds to a sudden increase in the interlayer exciton energy when increasing its density, i.e., an incompressible state of the interlayer exciton. Moreover, this state connects smoothly into the ν_e= 1 electron correlated insulator, indicating one particle per moiré site along the entire path and ν_{ex}= 1 in the limit of sole occupation by excitons [we have also independently calibrated the exciton density (*34*)]. Our observation of an incompressible exciton state at ν_{ex}= 1 (half-filling of an exciton band considering the valley degeneracy) and charge neutrality is a hallmark of a correlated insulator purely made of excitons. This correlated insulator could be a bosonic Mott insulator (*28*, *37*), a generalized Wigner crystal (*38*, *39*), or a charge transfer insulator (*40*). We unambiguously exclude the charge transfer insulator scenario through valley (flavor)–resolved measurements (*34*). Estimations of r_s (ratio of Coulomb interaction to kinetic energy) and the Mott criterion suggest

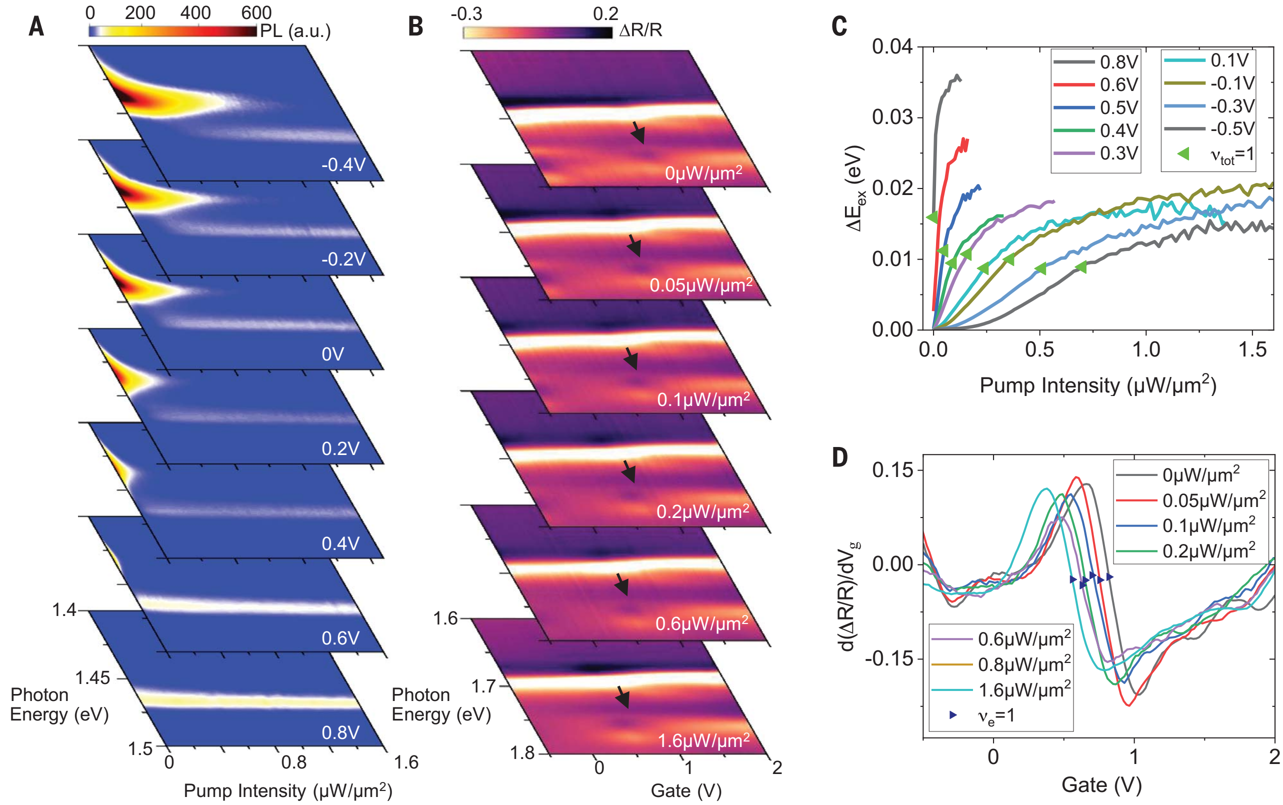


Fig. 2. Mixed correlated insulator. (**A**) Pump intensity–dependent PL spectra at gate voltages V_g from −0.4 V (near charge neutrality) to 0.8 V (electron-one-filling). (**B**) Gate-dependent absorption spectra at pump intensities from 0 to 1.6 μW/μm^2. Blue arrows mark the kink and splitting in absorption peaks at $\nu_e = 1$. (**C**) Power-dependent interlayer exciton energy change ΔE_{ex} at representative gate voltages. Green triangles mark middles of the transitions, which appear at smaller pump intensity (ν_{ex}) with increasing gate voltage (ν_e), consistent with a mixed correlated insulator state at $\nu_{tot} = 1$. (**D**) First-order derivative of absorption spectra with respect to gate voltage at 1.76 eV under different pump intensities. Blue triangles denote $\nu_e = 1$. All measurements are performed at a base temperature of 1.65 K in device D1.

a bosonic Mott insulator nature of our observation (*34*, *38*, *39*).

Mixed correlated insulator

To quantify the jump in interlayer exciton energy, we fit PL spectra with two Lorentzian peaks (*34*) and compute the exciton energy change $\Delta E_{ex} = [(E_1 I_1 + E_2 I_2)/(I_1 + I_2) - E_1]$, which can be considered as an effective exciton chemical potential. Here, E_1 (E_2) and I_1 (I_2) correspond to the energy and amplitude of the PL peak I and II, respectively (Fig. 1D). Figure 2C summarizes the evolution of ΔE_{ex} with pump intensity at representative gating, where a transition in exciton energy is clearly observed at all gate voltages. The transition appears not particularly sharp around charge neutrality, largely because of the nonlinear dependence of exciton density on pump intensity (Fig. 1D). In addition, the transition is broadened by spatial inhomogeneity in the exciton density that is expected to be much larger than the charge case (*34*). At charge neutrality $V_g = -0.5$ V, we determine the position of $\nu_{ex} = 1$ from the middle point of the ΔE_{ex} transition. We also independently determine the exciton density at this point to be $(2 \pm 0.2) \times 10^{12}$ cm^{-2} from time-resolved measurements (*34*), which matches well with the expected density of moiré cells $n_0 = 2.1 \times 10^{12}$ cm^{-2}. At finite electron density, the transition happens when electrons and excitons cooperate to occupy all available sites. The transition points therefore correspond to $\nu_{tot} = \nu_e + \nu_{ex} = 1$ (green triangles in Fig. 2C) and keep shifting to a lower pump intensity until reaching $\nu_{ex} = 0$ when $\nu_e = 1$.

To separate ν_e and ν_{ex} better, we also measure absorption spectra of the moiré superlattice while varying gate voltage and pump intensity. Because intralayer excitons are only sensitive to ν_e but not ν_{ex} (Fig. 1, E and F), we use spectral changes in absorption to independently determine ν_e. Figure 2B shows the gate-dependent absorption spectra of the moiré bilayer at different pump intensities. All plots show similar behaviors except for a slight shift in gate voltage. This can be seen more clearly in the gate-differentiated absorption spectra at representative pump intensity (Fig. 2D) (*34*). Blue triangles in Fig. 2, B and D, label the position of "kinks" that correspond to $\nu_e = 1$.

With both ν_{tot} and ν_e extracted, Fig. 3A summarizes the phase diagram spanned by ν_e and ν_{ex} at a base temperature of 1.65 K. The color scale represents ΔE_{ex}; the boundaries of $\nu_{tot} = 1$ and $\nu_e = 1$ are overlaid on the phase diagrams. The $\nu_e = 1$ line shifts slightly toward lower gate voltage at high pump intensity, presumably from photocarriers generated during the relaxation of excitons. However, this effect is rather weak, and the $\nu_e = 1$ line deviates far from the $\nu_{tot} = 1$ line where transitions in ΔE_{ex} are observed. This further confirms that the transition at $\nu_{tot} = 1$ is not from an electron lattice. Instead, it originates from a mixed lattice composed of both excitons and electrons. Our observations, therefore, suggest a mixed correlated insulator along the $\nu_{tot} = 1$ line, which smoothly connects into a bosonic (fermionic) correlated insulator at the end point of $\nu_{ex} = 1$ ($\nu_e = 1$).

We further obtain the phase diagrams at 15 and 30 K (Fig. 3, B and C; see fig. S9 to S12 for details). A sharp transition in ΔE_{ex} is always observed at $\nu_e = 1$, consistent with the strong charge correlation and a high melting temperature of charge correlated insulator in WSe_2/WS_2 moiré superlattices (*3*, *4*). Similarly, the behavior at $\nu_{ex} = 1$ remains largely unchanged with temperature, indicating survival of the bosonic correlated insulator to above 30 K. By

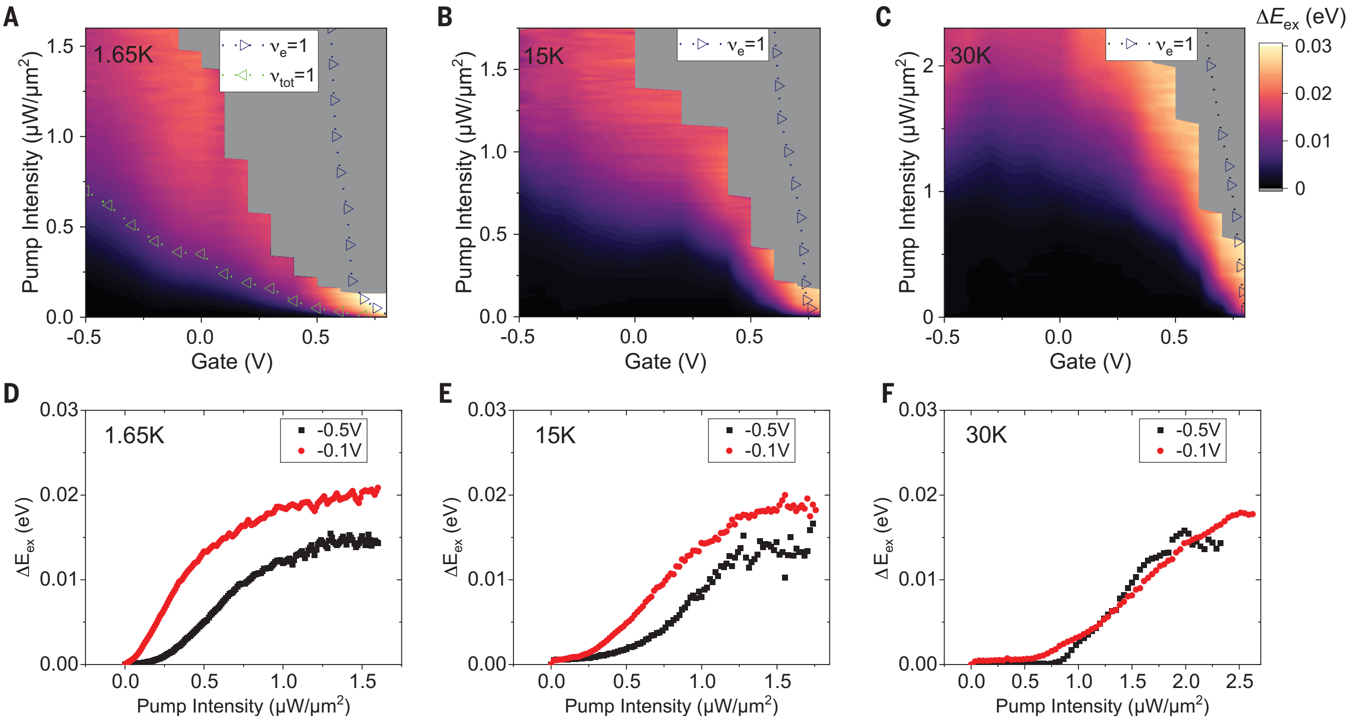


Fig. 3. Phase diagram. (**A** to **C**) ΔE_{ex} with respect to the gate voltage and pump intensity at 1.65 K (A), 15 K (B), and 30 K (C) measured on device D1. The boundaries of $\nu_{tot} = 1$ and $\nu_e = 1$, as determined from the pump-probe PL and absorption measurements, respectively, are highlighted with green and blue triangles. The clear separation between the two boundaries confirms that the transition of ΔE_{ex} at $\nu_{tot} = 1$ is not from a charge correlated insulator state until close to $\nu_e = 1$. Regions of high pump intensity and/or high gating are not shown because peak I has already disappeared and cannot be fitted reliably (*34*). (**D** to **F**) Vertical line-cuts of the phase diagrams for $V_g = -0.5$ V (charge neutrality) and −0.1 V (electron doped) at 1.65 K (D), 15 K (E), and 30 K (F). Whereas at 1.65 K, ΔE_{ex} rises faster for $V_g = -0.1$ V than −0.5 V, at 30 K it rises slower for $V_g = -0.1$ V than −0.5 V, suggesting partial melting and lower stability of the mixed correlated insulator state.

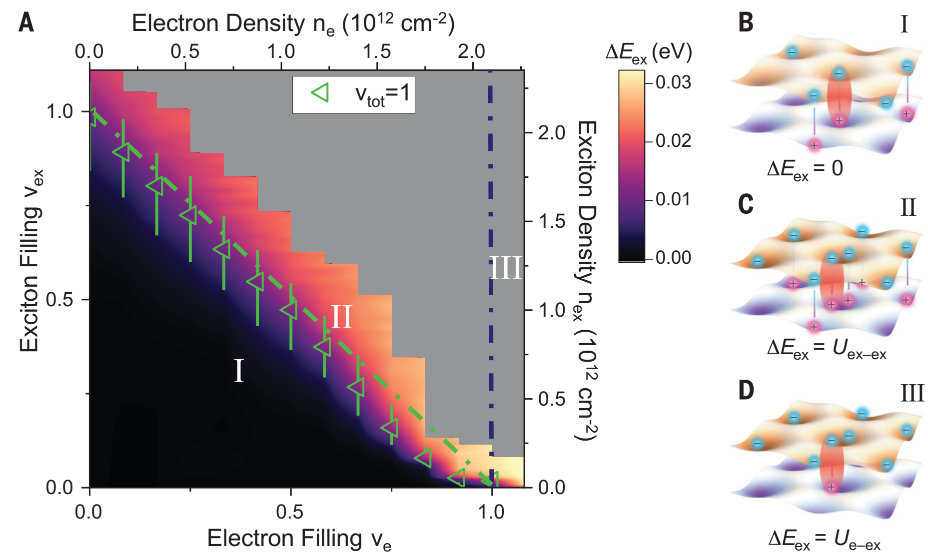


Fig. 4. Mixed Hubbard model. (**A**) ΔE_{ex} phase diagram measured on another 60°-aligned moiré bilayer (device D2) at 1.65 K with calibrated ν_{ex} and ν_e. Green ($\nu_{tot} = 1$) and blue ($\nu_e = 1$) dashed lines are boundaries predicted by a mixed Hubbard model, where ΔE_{ex} is expected to jump. Green triangles label the experimental 50% transition point in ΔE_{ex}, which matches well with the prediction. Error bars denote the range between 20 and 80% transition in ΔE_{ex}. (**B** to **D**) Schematics showing the energy required to add one interlayer exciton (red) into the system for region I (B), II (C), and III (D) in the phase diagram. Additional energy cost of $\Delta E_{ex} = U_{ex-ex}$ and U_{e-ex} is required in region II and III, respectively, from on-site repulsion between excitons and between electron-exciton.

contrast, for the regions in between, the transition in ΔE_{ex} becomes much slower than at base temperature. This can be seen clearly by comparing the ΔE_{ex} evolution at charge neutral ($V_g = -0.5$ V) and finite electron density ($V_g = -0.1$ V) (Fig. 3, D to F). At base temperature the finite ν_e case shows a sharper rise than charge neutral (Fig. 3D), which can be naturally explained by the smaller ν_{ex} needed to reach $\nu_{tot} = 1$ and therefore less inhomogeneous broadening from exciton density. Notably, at 30 K the rise in ΔE_{ex} becomes smoother at $V_g = -0.1$ V compared to charge neutral, indicating increasing exciton compressibility and partial melting of the mixed correlated insulator. This observation suggests that the mixed correlated insulator is less stable than both components (*34*).

Discussion

To allow direct comparison with theory, we calibrate ν_{ex} throughout the phase diagram by

time-resolved measurements (*34*) (Fig. 4A). The ΔE_{ex} jumps (green triangles) indeed occur at $\nu_{ex} + \nu_e = 1$ (green dashed line), further supporting the mixed correlated insulator origin. The experimental phase diagram matches very well with predictions from a two-component Hubbard model with both fermionic and bosonic species (*34*). Along the horizontal axis ($\nu_{ex} = 0$), ΔE_{ex} shows a sudden jump at $\nu_e = 1$. This can be understood because excitons always avoid electron-occupied sites owing to the on-site electron-exciton repulsion [Fig. 4B; see discussions in (*34*)] until all sites are electron-occupied at $\nu_e = 1$, after which adding an exciton necessarily pays the additional energy cost of $U_{e\text{-}ex}$ (Fig. 4D). We thereby determine $U_{e\text{-}ex}$ to be 32 to 40 meV (varies between samples). Such doping dependence is distinctively different from that in monolayer TMDC or TMDC bilayers where strong correlation is not observed, such as $WSe_2/MoSe_2$ (*20*, *21*). In these systems, electron doping immediately leads to the formation of trion and a corresponding emission peak in PL at lower energy (*20*, *21*). In WSe_2/WS_2, by contrast, the emission spectrum remains largely unchanged until $\nu_e = 1$, after which the PL shows a blueshift of ~35 meV and ~20 meV in 60° and 0° twisted moiré superlattices, respectively (*6*, *8*). Our model provides a natural explanation to this widely observed yet not fully understood PL doping dependence.

Similarly, ΔE_{ex} remains 0 along the vertical axis until $\nu_{ex} = 1$, after which adding an exciton requires an additional energy cost of $U_{ex\text{-}ex}$ (Fig. 4C). We thereby determine $U_{ex\text{-}ex}$ to be 14 to 20 meV (varies between samples). Away from the axes, ΔE_{ex} remains 0 for $\nu_{ex} + \nu_e < 1$ (region I in Fig. 4A) because added excitons can find an empty site to avoid on-site repulsion. For $\nu_{ex} + \nu_e \geq 1$ but $\nu_e < 1$ (region II), added excitons will prefer to stay on an exciton site because $U_{ex\text{-}ex} < U_{e\text{-}ex}$, and the additional energy cost is $\Delta E_{ex} = U_{ex\text{-}ex}$. For $\nu_e \geq 1$ (region III), all sites are already occupied by electrons and added excitons can only reside on an electron site; therefore $\Delta E_{ex} = U_{e\text{-}ex}$.

The above analysis ignores the hopping term or the two pseudospins of excitons from the K and K′ valleys (*25*). Nevertheless, it captures all salient features of the experiments. Several interesting theoretical questions remain to fully understand the phase diagram, such as on the effects of finite hopping, as well as the microscopic mechanism of exciton-exciton interaction and its dependence on exciton pseudospin, for which our results provide a valuable experimental reference.

REFERENCES AND NOTES

1. D. M. Kennes *et al*., *Nat. Phys.* **17**, 155–163 (2021).
2. Y. Cao *et al*., *Nature* **556**, 43–50 (2018).
3. E. C. Regan *et al*., *Nature* **579**, 359–363 (2020).
4. Y. Tang *et al*., *Nature* **579**, 353–358 (2020).
5. Y. Shimazaki *et al*., *Nature* **580**, 472–477 (2020).
6. S. Miao *et al*., *Nat. Commun.* **12**, 3608 (2021).
7. X. Huang *et al*., *Nat. Phys.* **17**, 715–719 (2021).
8. E. Liu *et al*., *Phys. Rev. Lett.* **127**, 037402 (2021).
9. C. Jin *et al*., *Nat. Mater.* **20**, 940–944 (2021).
10. A. Ghiotto *et al*., *Nature* **597**, 345–349 (2021).
11. T. Li *et al*., *Nature* **597**, 350–354 (2021).
12. T. Li *et al*., *Nature* **600**, 641–646 (2021).
13. C. Jin *et al*., *Nature* **567**, 76–80 (2019).
14. K. L. Seyler *et al*., *Nature* **567**, 66–70 (2019).
15. K. Tran *et al*., *Nature* **567**, 71–75 (2019).
16. E. M. Alexeev *et al*., *Nature* **567**, 81–86 (2019).
17. L. Zhang *et al*., *Nature* **591**, 61–65 (2021).
18. H. Yu, G.-B. Liu, J. Tang, X. Xu, W. Yao, *Sci. Adv.* **3**, e1701696 (2017).
19. F. Wu, T. Lovorn, A. H. MacDonald, *Phys. Rev. Lett.* **118**, 147401 (2017).
20. X. Wang *et al*., *Nat. Nanotechnol.* **16**, 1208–1213 (2021).
21. E. Liu *et al*., *Nature* **594**, 46–50 (2021).
22. X. Wang *et al*., *Nature* **604**, 468–473 (2022).
23. J. Gu *et al*., *Nat. Phys.* **18**, 395–400 (2022).
24. Z. Zhang *et al*., *Nat. Phys.* **18**, 1214–1220 (2022).
25. P. Rivera *et al*., *Nat. Nanotechnol.* **13**, 1004–1015 (2018).
26. S. Sugawa *et al*., *Nat. Phys.* **7**, 642–648 (2011).
27. C. Lagoin, S. Suffit, K. Baldwin, L. Pfeiffer, F. Dubin, *Nat. Phys.* **18**, 149–153 (2022).
28. M. P. A. Fisher, P. B. Weichman, G. Grinstein, D. S. Fisher, *Phys. Rev. B Condens. Matter* **40**, 546–570 (1989).
29. F. P. Laussy, A. V. Kavokin, I. A. Shelykh, *Phys. Rev. Lett.* **104**, 106402 (2010).
30. P. Rivera *et al*., *Nat. Commun.* **6**, 6242 (2015).
31. W. Li, X. Lu, S. Dubey, L. Devenica, A. Srivastava, *Nat. Mater.* **19**, 624–629 (2020).
32. Y. Cao *et al*., *Nature* **556**, 80–84 (2018).
33. S. Luryi, *Appl. Phys. Lett.* **52**, 501–503 (1988).
34. See additional text and data in the supplementary materials.
35. P. Nagler *et al*., *Mater* **4**, 025112 (2017).
36. Z. Wang, Y. H. Chiu, K. Honz, K. F. Mak, J. Shan, *Nano Lett.* **18**, 137–143 (2018).
37. M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, I. Bloch, *Nature* **415**, 39–44 (2002).
38. B. Padhi, C. Setty, P. W. Phillips, *Nano Lett.* **18**, 6175–6180 (2018).
39. B. Padhi, R. Chitra, P. W. Phillips, *Phys. Rev. B* **103**, 125146 (2021).
40. Y. Zhang, N. F. Q. Yuan, L. Fu, *Phys. Rev. B* **102**, 201115 (2020).
41. R. Xiong, Data repository for: Correlated insulator of excitons in WSe2/WS2 moire superlattices, Open Science Framework (2023).

ACKNOWLEDGMENTS

R.X. thanks T. Xie, C. Tschirhart, A. Potts, Y. Guo, W. Gong, and W. Y. Xing for experimental help. **Funding:** C.J. acknowledges support from Air Force Office of Scientific Research under award FA9550-23-1-0117. R.X. acknowledges support from the UC Santa Barbara NSF Quantum Foundry funded via the Q-AMASE-i program under award DMR-1906325. S.T. acknowledges support from DOE-SC0020653 (materials synthesis), NSF DMR-2206987 (elemental purification), NSF ECCS 2052527 (electrical characterization for crystal optimization), DMR 2111812 (excitonic characterization during growth optimization), and CMMI 2129412 (large size crystal development). K.W. and T.T. acknowledge support from the JSPS KAKENHI (grant nos. 19H05790, 20H00354, and 21H05233). **Author contributions:** C.J. conceived and supervised the project. R.X. and J.H.N. fabricated the devices. R.X. and S.L.B performed the optical measurements and analyzed the data. P.H., R.S., and S.T. grew the WSe_2 and WS_2 crystals. K.W. and T.T. grew the hBN crystals. C.J. and R.X. wrote the manuscript with the input from all the authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data in the main text and supplementary materials, as well as the code for peak fitting, are available from Open Science Framework (*41*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add5574
Materials and Methods
Supplementary Text
Figs. S1 to S16
References (*42*–*63*)

Submitted 24 June 2022; resubmitted 26 September 2022
Accepted 27 April 2023
Published online 11 May 2023
10.1126/science.add5574

WORKING LIFE

By **Emma Xu**

How I caught up

A few weeks into my Ph.D. program, as the new students were still getting acquainted with one another, the conversation turned to our parents' professions. I wasn't surprised to learn that quite a few of my new classmates had parents who were professors; the rest were in other highly respected professions. I was too embarrassed to share that my parents had no higher education. They were denied it because they grew up during the Cultural Revolution in China. As a result, I was late learning about science; I didn't even know what research was until I was 20 years old. For much of my scientific training I feared I was behind my peers. But I have been able to close the gap—and that is something to be proud of.

When I was 15, my mother and I immigrated with my American stepfather to a small town in Texas. My English was insufficient even for simple conversations, and I struggled both socially and academically. We lived below the poverty line and my stepfather subjected my mother and me to horrible abuse, leading to my mother being hospitalized with severe depression when I was 18. Not long after, I escaped from our dingy apartment with a suitcase of essentials and never looked back.

In the months that followed, I spent my days studying full-time at the small local college and working various part-time jobs. At night I couch surfed at friends' places or slept in my car in parking lots. I constantly worried about my next meal and where I was going to spend the night. I was just trying to survive, day by day, with no prospects for the future.

But I finally had control over my life. A year after leaving, I managed to transfer to a larger university in a city about 2 hours away, where my mother joined me after she recovered. With the newfound freedom and mental space to explore my academic interests, I became intrigued by physics and the possibility of contributing to understanding how the universe works.

During a department social, someone told me about a thing called "research." I didn't know what that entailed, but I figured it might be a good thing to try. I contacted every professor in the department asking for a research position until one took me on. At age 20, I finally got a taste of research. I was fascinated to see the theories I learned in the classroom come to life. Interacting with graduate students and postdocs, who seemed so smart and knowledgeable, inspired my own desire to pursue a career in research.



"A desire to learn [is] at least as valuable as childhood advantages."

But I frequently felt insecure about my prospects. One physics professor liked to show off photos of his son as a boy with Nobel laureates and brag about how the son started doing research as a high schooler. Oh, yes, his wife was a physics professor, too. I feared my late start meant my scientific career was doomed to failure.

Still, the joy and fulfillment I derived from research quieted my negative thinking, and I focused on improving myself. Over the remainder of my undergraduate years, I seized every opportunity to gain research experience at various institutions and in a range of disciplines. With each new skill I acquired, my confidence grew. After a gap year working at research-oriented startups, I felt ready for graduate school.

There were times early on when I still felt behind and intimidated, like that early conversation with my classmates about our parents. But reflecting on how I got here, I came to realize that a desire to learn and willingness to put in the necessary effort are at least as valuable as childhood advantages.

I'm now 6 years into my Ph.D. program and proud of what I've accomplished, personally and academically. I still encounter situations where it seems that my peers are more advanced than I am. But I no longer feel insecure. Instead, I am now equipped with the confidence to tackle my weaknesses head-on and celebrate the strengths that set me apart. ■

Emma Xu is a Ph.D. student at Columbia University. Send your career story to SciCareerEditor@aaas.org.



Bill Ochs, John Mather, Mani L. Bhaumik, and Major General Charles Frank Bolden Jr. attend the 3 May award ceremony.

AAAS bestows new award

Inaugural Mani L. Bhaumik Breakthrough of the Year Award honors NASA contributors

By **Meagan Phelan**

The American Association for the Advancement of Science's inaugural Mani L. Bhaumik Breakthrough of the Year Award—focused on standout contributors to NASA's James Webb Space Telescope—honors three individuals who supported vast swaths of the JWST community over decades and whose persistence amid multiple setbacks ensured the mission's completion.

The award recognizes Major General Charles Frank Bolden Jr., USMC (Ret.), a former administrator of the National Aeronautics and Space Administration; John Mather, senior project scientist of the JWST since 1995; and Bill Ochs, JWST project manager from 2011 through the telescope's launch.

The award selection committee seeks to acknowledge not only the winners' individual contributions but also the teams they inspired, whose collective work has given us all a completely different view of the Universe. Recognizing such significant scientific developments, and the people behind them, is a central philosophy for donor Mani L. Bhaumik—a physicist with myriad contributions to the development of high-powered lasers—as well as for AAAS and the *Science* family of journals.

"The JWST's ability to go back in time to help us see the earliest objects in the Universe will dramatically increase our view of the cosmos. This is merely one of its superb capabilities. I'm thrilled to see this award go to individuals whose persistence on such a difficult but worthy scientific endeavor made it a reality," said Bhaumik.

The Bhaumik Breakthrough of the Year Award builds upon the well-established *Science* Breakthrough of the Year, the journal's choice of the top research advance of the year. In 2022, the deployment of the JWST, which first beamed images and data back to Earth in July, earned Breakthrough status. The telescope, the product of a 20-year collaboration led by NASA and with the partnership of the European and Canadian space agencies, was launched on 25 December 2021.

To select just three winners whose work best exemplified the journey to bring JWST to life, *Science* journals Editor-in-Chief Holden Thorp convened a committee.

"So many people have touched the JWST project," said committee member Ellen Stofan, under secretary for science and research at the Smithsonian and former chief scientist at NASA. "This global collaboration exemplifies what a NASA mission is all about."

But the sheer number of contributors to the decades-long JWST project also made selecting three winners an incredible challenge.

"We felt that we needed to represent the teamwork and various skillsets necessary to make this mission a success," said Stofan.

The committee agreed that they wanted the inaugural award to represent three key groups: the scientists, the engineers, and the leaders who advocated for the JWST mission in Congress when its status was under great threat.

"Our focus was to think less about the individual contributions and more about how an individual represented a key part of the team's success," said committee member Michael Moloney, CEO of

the American Institute of Physics. “That isn’t to say that any of these individuals didn’t make enormous individual contributions, but that they did that in a broader context—representing the work of literally thousands of other people that led to the science we now see being done, and that will be done.”

A common theme among all three winners was their persistence.

For Bolden, a retired United States Marine Corps Major General and a former astronaut who served as administrator of NASA from 2009 to 2017, a pivotal moment came early in his tenure at NASA, in 2010, when an internal government review panel flagged major budgetary problems for JWST.

“In the face of such a situation, the easy decision would have been to cancel the mission,” said Stofan.

But Bolden—passionate about the telescope, with a deep understanding of its potential for scientific discovery—responded by elevating the telescope’s completion to an agency-wide priority and ensuring that it remained one.

“He invested a great deal of his personal time and capital in working with key partners in Congress, at NASA, and beyond to maintain the agency’s support for the project,” said Moloney.

Reflecting on this time, Bolden said that the importance of advocating for JWST never wavered in his mind, even when he was up against strong challengers deeply concerned about cost and time.

“This telescope was going to revolutionize our understanding of astrophysics,” said Bolden, whose enthusiasm for JWST came in part from his time as copilot crewmember of the mission that deployed the Hubble Space Telescope in 1990. “What we’d learn from JWST would dwarf what we’d learned from Hubble.”

During Bolden’s leadership at NASA—from 2009 to 2017—thousands of people were key contributors to the JWST mission’s ultimate launch and deployment.

“As a winner of this award, Bolden’s representing the whole agency because so many of the key milestones in JWST’s development happened during his tenure,” said Moloney.

Among these milestones were scientific advances overseen by Mather, the telescope’s senior project scientist, who played a key role in the $10 billion project from the earliest days.

“He represents scientists who had a dream about building a telescope like this even before Hubble was launched,” said Moloney.

In his work leading the JWST science team beginning in 1995, Mather represented scientific interests to project management and helped see each step forward, no matter how small.

“We set ourselves an incredible challenge,” said Mather. “The telescope represented a very revolutionary idea, and we had to invent many things.”

His work included setting up international science contests to find the best talent to invent the right instruments and technologies for the telescope, and to bring the same inventions to life.

“John was able to consistently articulate why it was so worth spending money on the JWST mission,” said Stofan. “The promise of the science, which he articulated so well, kept people going.”

“I didn’t fear we would lose support because there was no alternative to what we were doing,” said Mather. Mather is known for speaking passionately about the promise of the telescope for making observations of events and objects at sufficient distances to reveal the Universe when it was less than a billion years old.

“John represents the hopes and dreams of more than a generation of scientists, and now a new generation of scientists who are working on the mission, too,” said Moloney.



For Ochs, who served as the JWST project manager starting in 2011—representing both engineering and science teams—coming onboard when he did meant confronting tough realities: An external review had just concluded that the JWST wasn’t going to make its launch date.

“I had to learn this huge, complex mission and figure out a new launch estimate and cost estimate,” said Ochs.

He was very focused on the people at the center of the project.

“When I first got there, morale was down,” Ochs said. He described his approach to management—focused on efforts to bring people together to feel heard and seen. He also preferred walking around to visit people doing the work, instead of waiting for them to come to him. He believed in being a strong listener, managing with humor and building consensus.

“This mission required too many miracles,” said Stofan. “Bill knew how to listen, how to make tough decisions, and how to get the team to work together to reach the finish line.”

Among the miracles he oversaw was recovery from an incident in 2016 when the telescope’s sunshield, a honeycomb array of highly smooth mirrors that unfolds in space, was evaluated in a shake test designed to mimic the rocket launch environment. At first, results looked good, but after opening the door of the shake test chamber, “there were screws, nuts and washers lying all over the floor,” said Ochs. This put the team back 6 months.

Moving forward from that incident, Ochs reflected on the importance of building a work culture in which people didn’t feel burdened by a deadline. He would visit people every shift, encouraging them to focus on accuracy instead of time pressure. “If you see something small, tell us now,” he’d say. “I’d rather lose 30 minutes now than 6 months later.”

While the pandemic ultimately slowed the engineering and science teams, the telescope was successfully launched in 2021.

“To get to that point required really robust project management,” said Moloney. “Bill was there to orchestrate all of it and make sure the orchestra was playing for the launch on Christmas Day [2021].”

“It was the best job of my career,” Ochs said.

On 3 May, a live event at AAAS headquarters in Washington, D.C., honored Ochs, Mather, and Bolden. Each winner spoke about their time serving the JWST community.

The cash prize each winner receives—the largest bestowed by any scientific society—is made possible by the biggest transformational gift in AAAS’s history. The $11.4 million pledge to the organization supports a $250,000 cash prize annually.

“To be able to recognize hard-working scientists with this award—and perhaps prompt even more good work—is exciting,” said donor Mani Bhaumik in a Q&A with AAAS News & Notes last July.

Stofan said that compared to other science prizes, the Bhaumik Breakthrough of the Year Award stands poised to fill an important gap. “The problems of the future are interdisciplinary problems. They cannot be solved by one person. The way that this prize has unfolded this year—the fact that we’ve selected three people with different skillsets and who came in at different stages of the mission—shows you that you need teams and various skills and approaches to tackle problems of the future.”

“This award will help elevate scientists working in research areas most likely to benefit society whose approaches to both leadership and problem-solving represent the way forward,” said Thorp. “We hope this will inspire future team-oriented approaches to tackling major problems the world faces.”